



**UNIVERSIDAD POLITECNICA DE MADRID
FACULTAD DE INFORMATICA**

TESIS DOCTORAL

**APORTACION AL ANALISIS DE
ESPECTROS DE RADIACION.
ESTRUCTURAS Y ALGORITMOS**

Autor: JUAN CARLOS DIAZ MARTIN

Directores: PEDRO GOMEZ VILDA

VICTORIA RODELLAR BIARGE

Madrid, 1994

TESIS DOCTORAL

APORTACION AL ANALISIS DE ESPECTROS DE RADIACION. ESTRUCTURAS Y ALGORITMOS

Presentada en la

FACULTAD DE INFORMATICA

de la

UNIVERSIDAD POLITECNICA DE MADRID

para la obtención del grado de

DOCTOR EN INFORMATICA

AUTOR: **JUAN CARLOS DIAZ MARTIN**

DIRECTORES: **PEDRO GOMEZ VILDA
VICTORIA RODELLAR BIARGE**

Madrid, 1994

a Marce y Jacinto,
mis padres

Indice

Resumen		v
Abstract		ix
Motivaciones		xiii
Objetivos		xvii
Capítulo 1	<i>Introducción a la detección y medida de radiación</i>	1
1.1	Naturaleza estadística de la emisión radiactiva	2
1.2	Interacción de los rayos X y gamma con la materia	3
1.3	Introducción a los detectores de radiación	6
1.3.1	Conceptos básicos	8
1.4	Detectores de gas	10
1.4.1	Relación entre tensión aplicada y carga recogida	11
1.4.2	Tipos de contadores de gas	13
1.5	Detectores de centelleo	14
1.6	Detectores de semiconductor	17
1.6.1	Operación de un detector semiconductor	17
1.6.2	Operación de la unión P-N como detector	20
1.6.3	Distintos tipos de semiconductor	24

1.6.3.1	Detectores de difusión de unión	24
1.6.3.2	Detectores de silicio compensados con litio	25
1.6.3.3	Detectores de germanio compensados con litio	26
1.6.3.4	El yoduro de mercurio y el telurio de cadmio	27
1.7	Instrumentación básica de medida de radiación	28
1.7.1	Polarización del detector	28
1.7.2	El preamplificador	29
1.7.3	El amplificador	29
1.7.4	Conversores analógico-digitales	31
1.7.5	El analizador multicanal	31
1.7.6	Calibración de un analizador multicanal	33
1.8	Bibliografía	34
Capítulo 2	<i>El análisis de datos en espectroscopia gamma</i>	37
2.1	La función de transferencia del detector	38
2.1.1	El espectro proporcionado por el detector	39
2.1.2	El espectro observado en el analizador multicanal	41
2.2	La deconvolución del espectro observado	42
2.3	Mejora de resolución del espectro medido	46
2.4	Búsqueda de picos	49
2.4.1	Descripción funcional del pico	50
2.4.2	La detección del pico	52
2.4.3	Técnicas de ajuste	54
2.4.4	Tiempos y recursos de computación	55
2.5	Técnicas de regresión multilineal	56
2.5.1	Formación del espectro. Conceptos fundamentales	56
2.5.2	El modelo de regresión lineal múltiple	58
2.5.3	El modelo estadístico del espectro problema	59
2.5.3.1	Bondad del encaje	61
2.6	Bibliografía	62
Capítulo 3	<i>La descomposición espectral vía memoria asociativa</i>	73
3.1	Definición del problema	74
3.2	Introducción a la teoría de pseudoinversas	77

3.2.1	Pseudoinversa y mínimos cuadrados	77
3.2.2	El algoritmo de Greville	80
3.3	Memorias asociativas lineales	82
3.3.1	La memoria lineal óptima	84
3.3.2	Computación recursiva de memorias lineales	85
3.3.2.1	El algoritmo de proyección del gradiente	86
3.3.2.2	El algoritmo de Widrow-Hoff	87
3.4	Descomposición espectral vía memoria asociativa	91
3.4.1	Derivación de la memoria	91
3.4.2	Especificación de algoritmos	92
3.4.2.1	El algoritmo de proyección del gradiente	93
3.4.2.2	El algoritmo de Widrow-Hoff	95
3.5	Implementación <i>VLSI</i> de la memoria asociativa	96
3.6	Criterios de estimación del error	99
3.6.1	Validación del modelo de formación	100
3.6.2	Estimación el error del vector de contribuciones	100
3.7	Buscando el operador mínimo	104
3.8	Bibliografía	108
Capítulo 4	<i>Ampliación de la memoria ante pequeñas derivas instrumentales</i>	113
4.1	Descripción analítica del espectro con distorsión instrumental	114
4.2	Memoria asociativa y pequeñas derivas instrumentales	118
4.2.1	Derivas en espectros puros	122
4.2.2	Derivas en mezclas	126
4.2.3	Discusión en torno a la interpretación de derivas	131
4.2.3.1	Deriva en umbral	132
4.2.3.2	Deriva en ganancia	132
4.3	El orden de la expansión de Taylor	133
4.3.1	Desarrollos superiores al primero	134
4.3.2	Reducción de datos	138
4.3.3	Cálculo numérico de los desarrollos	140
4.3.4	El espectro de fondo en detectores gamma	141
4.4	Buscando el operador mínimo	142
4.5	Bibliografía	144

Capítulo 5	<i>Derivas severas: memoria asociativa con realimentación</i>	147
5.1	Optimización cuadrática no lineal	148
5.1.1	Formulación general del problema de regresión no lineal	149
5.1.2	Revisión de métodos	150
5.1.2.1	El método del gradiente	151
5.1.2.2	El método de Gauss-Newton	152
5.1.2.3	El método de Newton	153
5.2	Criterios Metodológicos	156
5.3	El algoritmo de Gauss-Newton en el problema de la mezcla	157
5.3.1	Un problema de regresión no lineal	157
5.3.2	Una nueva formulación del problema	160
5.4	Memoria asociativa en el algoritmo de Gauss-Newton	162
5.4.1	Región de convergencia	166
5.5	Cómputo de las transformaciones no lineales	167
5.6	Bibliografía	169
Discusión de resultados		171
Conclusiones		177
Líneas futuras		181
Agradecimientos		185

Resumen

A lo largo del presente trabajo se investiga la viabilidad de la descomposición automática de espectros de radiación gamma por medio de algoritmos de resolución de sistemas de ecuaciones algebraicas lineales basados en técnicas de pseudoinversión. La determinación de dichos algoritmos ha sido realizada teniendo en cuenta su posible implementación sobre procesadores de propósito específico de baja complejidad.

En el primer capítulo se resumen las técnicas para la detección y medida de la radiación gamma que han servido de base para la confección de los espectros tratados en el trabajo. Se reexaminan los conceptos asociados con la naturaleza de la radiación electromagnética, así como los procesos físicos y el tratamiento electrónico que se hallan involucrados en su detección, poniendo de relieve la naturaleza intrínsecamente estadística del proceso de formación del espectro asociado como una clasificación del número de detecciones realizadas en función de la energía supuestamente continua asociada a las mismas. Para ello se aporta una breve descripción de los principales fenómenos de interacción de la radiación con la materia, que condicionan el proceso de detección y formación del espectro. El detector de radiación es considerado el elemento crítico del sistema de medida, puesto que condiciona fuertemente el proceso de detección. Por ello se

examinan los principales tipos de detectores, con especial hincapié en los detectores de tipo semiconductor, ya que son los más utilizados en la actualidad. Finalmente, se describen los subsistemas electrónicos fundamentales para el acondicionamiento y pretratamiento de la señal procedente del detector, a la que se le denomina con el término tradicionalmente utilizado de *Electrónica Nuclear*. En lo que concierne a la espectroscopía, el principal subsistema de interés para el presente trabajo es el *analizador multicanal*, el cual lleva a cabo el tratamiento cualitativo de la señal, y construye un histograma de intensidad de radiación en el margen de energías al que el detector es sensible. Este vector N -dimensional es lo que generalmente se conoce con el nombre de *espectro de radiación*. Los distintos radionúclidos que participan en una fuente de radiación no pura dejan su impronta en dicho espectro.

En el capítulo segundo se realiza una revisión exhaustiva de los métodos matemáticos en uso hasta el momento ideados para la identificación de los radionúclidos presentes en un espectro compuesto, así como para determinar sus actividades relativas. Uno de ellos es el denominado de *regresión lineal múltiple*, que se propone como la aproximación más apropiada a los condicionamientos y restricciones del problema: capacidad para tratar con espectros de baja resolución, ausencia del concurso de un operador humano (no supervisión), y posibilidad de ser soportado por algoritmos de baja complejidad capaces de ser instrumentados sobre procesadores dedicados de alta escala de integración.

El problema del análisis se plantea formalmente en el tercer capítulo siguiendo las pautas arriba mencionadas y se demuestra que el citado problema admite una solución en la *teoría de memorias asociativas lineales*. Un operador basado en este tipo de estructuras puede proporcionar la solución al problema de la descomposición espectral deseada. En el mismo contexto, se proponen un par de algoritmos adaptativos complementarios para la construcción del operador, que gozan de unas características aritméticas especialmente apropiadas para su instrumentación sobre *procesadores de alta escala de integración*. La característica de adaptatividad dota a la memoria asociativa de una gran flexibilidad en lo que se refiere a la incorporación de nueva información en forma progresiva.

En el capítulo cuarto se trata con un nuevo problema añadido, de índole altamente compleja. Es el del tratamiento de las deformaciones que introducen en el espectro las derivas instrumentales presentes en el dispositivo detector y en la electrónica de preacondicionamiento. Estas deformaciones invalidan el modelo de regresión lineal utilizado para describir el espectro problema. Se deriva entonces un modelo que incluya las citadas deformaciones como una ampliación de contribuciones en el espectro compuesto, el cual conlleva una ampliación sencilla de la memoria asociativa capaz de tolerar las derivas en la mezcla problema y de llevar a cabo un análisis robusto de contribuciones. El método de ampliación utilizado se basa en la suposición de pequeñas perturbaciones.

La práctica en el laboratorio demuestra que, en ocasiones, las derivas instrumentales pueden provocar distorsiones severas en el espectro que no pueden ser tratadas por el modelo anterior. Por ello, en el capítulo quinto se plantea el problema de medidas afectadas por fuertes derivas desde el punto de vista de la *teoría de optimización no lineal*. Esta reformulación lleva a la introducción de un algoritmo de tipo recursivo inspirado en el de Gauss-Newton que permite introducir el concepto de *memoria lineal realimentada*. Este operador ofrece una capacidad sensiblemente mejorada para la descomposición de mezclas con fuerte deriva sin la excesiva carga computacional que presentan los algoritmos clásicos de optimización no lineal.

El trabajo finaliza con una discusión de los resultados obtenidos en los tres principales niveles de estudio abordados, que se ofrecen en los capítulos tercero, cuarto y quinto, así como con la elevación a definitivas de las principales conclusiones derivadas del estudio y con el desglose de las posibles líneas de continuación del presente trabajo.

Abstract

Through the present research, the feasibility of Automatic Gamma-Radiation Spectral Decomposition by Linear Algebraic Equation-Solving Algorithms using Pseudo-Inverse Techniques is explored. The design of the before mentioned algorithms has been done having into account their possible implementation on Specific-Purpose Processors of Low Complexity. In the first chapter, the techniques for the detection and measurement of gamma radiation employed to construct the spectra being used throughout the research are reviewed. Similarly, the basic concepts related with the nature and properties of the hard electromagnetic radiation are also re-examined, together with the physic and electronic processes involved in the detection of such kind of radiation, with special emphasis in the intrinsic statistical nature of the spectrum build-up process, which is considered as a classification of the number of individual photon-detections as a function of the energy associated to each individual photon. For such, a brief description of the most important matter-energy interaction phenomena conditioning the detection and spectrum formation processes is given. The radiation detector is considered as the most critical element in the measurement system, as this device strongly conditions the detection process. For this reason, the characteristics of the most frequent detectors are re-examined, with special emphasis on those of semiconductor nature, as these are the most frequently employed ones nowadays. Finally, the fundamental electronic subsystems for pre-

aconditioning and treating of the signal delivered by the detector, classically addressed as *Nuclear Electronics*, is described. As far as Spectroscopy is concerned, the subsystem most interesting for the scope covered by the present research is the so-called *Multichannel Analyzer*, which is devoted to the qualitative treatment of the signal, building-up a histogram of radiation intensity in the range of energies in which the detector is sensitive. The resulting N -dimensional vector is generally known with the name of *Radiation Spectrum*. The different radio-nuclides contributing to the spectrum of a composite source will leave their fingerprint in the resulting spectrum. Through the second chapter, an exhaustive review of the mathematical methods devised to the present moment to identify the radio-nuclides present in the composite spectrum and to quantify their relative contributions, is reviewed. One of the more popular ones is the so-known *Multiple Linear Regression*, which is proposed as the best suited approach according to the constraints and restrictions present in the formulation of the problem, i.e., the need to treat low-resolution spectra, the absence of control by a human operator (un-supervision), and the possibility of being implemented as low-complexity algorithms amenable of being supported by *VLSI Specific Processors*. The analysis problem is formally stated through the third chapter, following the hints established in this context, and it is shown that the addressed problem may be satisfactorily solved under the point of view of *Linear Associative Memories*. An operator based on this kind of structures may provide the solution to the spectral decomposition problem posed. In the same context, a pair of complementary adaptive algorithms useful for the construction of the solving operator are proposed, which share certain special arithmetic characteristics that render them specially suitable for their implementation on *VLSI Processors*. The adaptive nature of the associative memory provides a high flexibility to this operator, in what refers to the progressive inclusion of new information to the knowledge base. Through the fourth chapter, this fact is treated together with a new problem to be considered, of a high interest but quite complex nature, as is the treatment of the deformations appearing in the spectrum when instrumental drifts in both the detecting device and the pre-aconditioning electronics are to be taken into account. These deformations render the Linear Regression Model proposed almost useless to describe the resulting spectrum. A new model including the drifts is derived as an extension of the individual contributions to the composite spectrum, which implies a simple extension of the Associative Memory, which renders this suitable to accept the drifts in the

composite spectrum, thus producing a robust analysis of contributions. The extension method is based on the *Low-Amplitude Perturbation Hypothesis*. Experimental practice shows that in certain cases the instrumental drifts may provoke severe distortions in the resulting spectrum, which can not be treated with the before-mentioned hypothesis. To cover also these less-frequent cases, through the fifth chapter, the problem involving strong drifts is treated under the point of view of *Non-Linear Optimization Techniques*. This reformulation carries the study to the consideration of recursive algorithms based on the *Gauss-Newton* methods, which allow the introduction of *Feed-Back Memories*, computing elements with a sensibly improved capability to decompose spectra affected by strong drifts. The research concludes with a discussion of the results obtained in the three main levels of study considered, which are presented in chapters third, fourth and fifth, together with the review of the main conclusions derived from the study and the outline of the main research lines opened by the present work.

Motivaciones

Los factores condicionantes de la elección de este tipo de línea de trabajo por el doctorando para la realización de la presente Tesis derivan de una serie de circunstancias, de entre las que cabe citar su incorporación como colaborador al Laboratorio de Tecnología de Computadores del Departamento de Arquitectura y Tecnología de Sistemas Informáticos de la U.P.M. para realizar su Trabajo Fin de Carrera en un tema de control robótico. Durante el período de realización de dicho trabajo, el doctorando se fue familiarizando con conceptos relacionados con la teoría de control y el tratamiento digital de la señal, así como con el aparato matemático subyacente a estas líneas. A finales de 1988, el citado Departamento firmó un convenio con el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), para el desarrollo de algoritmos de tratamiento de señal y reconocimiento de patrones que, utilizando técnicas adaptativas afines a las redes neurales, permitiesen al CIEMAT, en último extremo, desarrollar una cámara ligera de visión de radiación gamma. El doctorando fue propuesto por el Departamento de Arquitectura y Tecnología de Sistemas Informáticos como becario para el desarrollo de dicho convenio, y el CIEMAT lo adscribió al mismo. A partir de dicho momento, y hasta la finalización del convenio, el doctorando participó en el desarrollo de los algoritmos de descomposición y cuantificación de espectros de forma casi exclusiva. Fruto de los citados trabajos es la Tesis que se presenta a continuación.

Una cámara de radiación gamma se diferencia de una cámara convencional de vídeo en que opera en la región del espectro electromagnético correspondiente a la radiación gamma. El elemento sensible de esta cámara consiste en una matriz de $M \times M$ detectores de radiación. Cada uno de estos detectores mide la intensidad de radiación incidente en el mismo. Si identificamos cada detector por su posición (i, j) en la matriz, se obtiene lo que puede llamarse una imagen de intensidad de la radiación. Cabe la posibilidad de incorporar un analizador multicanal a cada detector de la matriz. En consecuencia, a cada pixel de la imagen se le asocia un vector N -dimensional de escalares, denominado *espectro de radiación*. El análisis de cada uno de estos espectros identifica los radio-núclidos que lo generan y determina, asimismo, la actividad correspondiente a cada uno de ellos. Este proceso desdobra la imagen de intensidad global de la fuente en las imágenes de intensidad de cada radio-núclido en particular. Esta ambiciosa capacidad de análisis era una de las prestaciones de que se quería dotar a la nueva cámara.

Otro de los requisitos de la citada cámara era su portabilidad, a fin de posibilitar su actuación en entornos radiactivos hostiles al operador humano. En principio, el principal obstáculo para conseguir tal portabilidad venía determinado por las características de los materiales detectores generalmente utilizados en radiación gamma, siendo éstos el yoduro de sodio compensado con talio, NaI(Tl), y el germanio compensado con litio, Ge(Li). Los primeros requieren un tamaño considerable, y los segundos un criostato de nitrógeno líquido, a fin de mantener el cristal a temperaturas de 77° K . La inviabilidad de estos detectores condujo a la elección del yoduro de mercurio (I_2Hg), semiconductor que opera a temperatura ambiente y cuyo comportamiento como detector de radiación gamma era prometedor, si bien muy poco conocido hasta entonces, sin aplicaciones comerciales, y aún objeto de investigación en laboratorio. Si el comportamiento del I_2Hg como detector en la región gamma fuese lo suficientemente bueno como para formar espectros con una resolución energética razonable, el camino hacia una cámara portable con posibilidad de construir imágenes selectivas a radio-núclidos quedaba abierto.

El problema se planteó entonces en términos de tener que aceptar como detector a un material que no generaría espectros de rayas, sino de tipo acampanado. Este tipo de espectros presentan el inconveniente de que si un elemento está presente en una mezcla en

una proporción relativamente baja, los efectos de sus picos en el espectro resultante quedan totalmente camuflados por la actividad de los elementos dominantes, imposibilitando una identificación por inspección visual del espectro, o por métodos basados en la detección de máximos en el mismo. Se imponía entonces la necesidad de desarrollar o adaptar algoritmos que pudiesen proceder a la identificación de espectros individuales por procedimientos automáticos no basados en el estudio de máximos, sino del espectro resultante en su conjunto. Los primeros estudios que se realizaron se llevaron a cabo utilizando un *modelo espectral Polo-Cero* basado en técnicas de *predicción lineal* que mostró de forma esperanzadora la posibilidad de resolver picos aparentemente invisibles en el espectro. Se tropezó entonces con la escasa compactabilidad de tales técnicas, y el hecho de que requiriesen la resolución de polinomios en z de alto orden, lo que las hacía poco atractivas. Se pasó por ello a considerar la posible candidatura de las *redes neurales* de tipo Perceptrón y Retropropagación, las cuales permitían una rápida identificación de los elementos presentes en las muestras sin precisar la localización de polos. Sin embargo, la naturaleza intrínsecamente no lineal de tales técnicas imposibilitaba el avanzar hacia un segundo escalón en el proceso de tratamiento del espectro, a saber, la cuantificación de las aportaciones relativas individuales presentes en el mismo. En este sentido, se inició el estudio que da origen al presente trabajo considerando la utilización de *memorias asociativas* basadas en la formación de la matriz pseudo-inversa del sistema lineal múltiple asociado, y fue en este caso cuando se comenzó a obtener resultados altamente satisfactorios. En una primera fase se consiguió identificar y cuantificar espectros compuestos en los que uno de los espectros estuviese incluso dos órdenes de magnitud por debajo de otro. Posteriormente se consideró la posibilidad de tratar los espectros deformados por derivas instrumentales débiles, y se consiguió extender el método de la matriz asociativa pseudo-inversa para resolver satisfactoriamente el mismo y, finalmente, se elaboraron otros tipos de metodologías afines para el tratamiento de derivas más serias.

Colateralmente, dado que la descomposición de $M \times M$ espectros N -dimensionales es un problema de alto orden, y que la rapidez en la algoritmia es un requisito fundamental, se consideró conveniente realizar un filtrado de las técnicas de *regresión lineal* utilizadas para considerar aquellas óptimamente adaptadas para su implementación sobre *procesadores específicos de alta escala de integración*. Las arquitecturas VLSI dedicadas constituyen una

de las líneas de investigación del Laboratorio de Tecnología de Computadores, adscrito al Departamento de Arquitectura y Tecnología de Sistemas Informáticos de la Universidad Politécnica de Madrid. En este sentido, el estudio para la realización práctica de tales estructuras, constituye una motivación añadida de considerable interés como línea de continuación del presente trabajo.

Objetivos

El trabajo desarrollado constituye un estudio sobre la aplicación y la posterior implementación VLSI del modelo estadístico de regresión múltiple, con el fin de resolver un problema concreto, que es el análisis de espectros de radiación gamma de baja resolución. La adopción de esta metodología de descomposición de espectro completo se sustenta sobre dos argumentos fundamentales. Primero, desde la perspectiva más amplia de la espectroscopía nuclear, la técnica alternativa de búsqueda de fotopicos en el espectro problema no resulta apropiada en espectros de baja resolución, debido a la dificultad o imposibilidad de discriminar entre picos anchos y solapados entre sí. Segundo, el estudio se enmarca dentro de un proyecto más amplio, cual es el de la construcción de una cámara de visión para radiación gamma, que incorpore la formación selectiva de la imagen de la radiación procedente de un radio-isótopo determinado. La consiguiente necesidad de rapidez en el análisis del espectro correspondiente a cada uno de los pixels de la imagen, parece desaconsejar la búsqueda de fotopicos, por constituir el ajuste de cada uno de ellos un problema adicional de regresión con varios parámetros no lineales. Recientes publicaciones sobre espectroscopía de absorción aconsejan asimismo la adopción de la vía del análisis del espectro completo.

Sentadas estas premisas, se pensó en iniciar un estudio metodológico del problema

empleando técnicas de regresión lineal múltiple, que permitiesen abarcar el estudio del espectro completo por procedimientos automáticos o semiautomáticos. Con este fin se fijaron los siguientes objetivos específicos:

1. Conseguir la descomposición de espectros de radiación generados por la superposición de la actividad de varios radio-núclidos puros mediante el marcado de la presencia o ausencia de los componentes individuales presentes en una base de espectros de referencia.
2. Cuantificación de las tasas de presencia de los mismos en el espectro compuesto indicando éstas términos relativos.
3. Abarcar bases de referencias que puedan contener hasta diez espectros referencia.
4. Permitir la construcción de algoritmos resolutores compactos y que puedan ser ampliables o modificables en forma recursiva.
5. Resolver espectros que contengan dos radio-núclidos con proporciones relativas de 1:100 como máximo.
6. Determinar un conjunto de algoritmos para identificación y cuantificación de espectros que sean suficientemente robustos ante distorsiones en el espectro compuesto producido por derivas en la ganancia y en el "offset" de la instrumentación del sensor.
7. Estudiar las posibles variantes algorítmicas que permitan la implementación de los mismos sobre procesadores específicos de alta escala de integración.

Capítulo *1*

INTRODUCCION A LA DETECCION Y MEDIDA DE RADIACION

El término radiación se usó hasta finales del siglo pasado para describir las ondas electromagnéticas. A la vuelta del siglo se descubrieron los rayos X, los electrones y la radiación natural. A pesar de que estos nuevos fenómenos mostraban características de partículas, también pasaron a formar parte del concepto de radiación. En los años 20, De Broglie desarrolló la teoría de la dualidad de la materia y puso fin a la controversia onda-partícula. Hoy en día se entiende por radiación todo el espectro electromagnético, así como las partículas atómicas y subatómicas que han sido descubiertas.

Podemos dividir la radiación en ionizante y no ionizante. Si una radiación es capaz de ionizar el medio que atraviesa se denomina ionizante. Radiación no ionizante son las ondas de radio, la luz visible y la luz ultravioleta. Radiación ionizante es el resto del espectro electromagnético de menor longitud de onda (rayos X y gamma). El material de este capítulo se refiere a esta última. Específicamente, se da una descripción, por una parte, de los principios físicos implicados en el proceso de interacción radiación-materia y, por otra, de la instrumentación básica de detección de radiación gamma que ha sido desarrollada en base al conocimiento de estos principios.

1.1 NATURALEZA ESTADISTICA DE LA EMISION RADIATIVA

La emisión de radiación es la cesión de energía de un sistema cuando pasa de un estado de mayor energía a otro de menor energía. De acuerdo con la teoría cuántica, esta energía es una cantidad discreta. Así, aunque se den las condiciones de transición de estado, esta puede producirse o no hacerlo. Con la introducción de la teoría cuántica, el destino de un sistema ya no está determinado por sus condiciones iniciales y la interacción con su entorno. El estudio de la Física cambió entonces de determinístico a probabilístico.

La emisión de radiación nuclear obedece las reglas de la Mecánica Cuántica y por tanto el número y la distribución energética de las partículas emitidas en un fenómeno nuclear tendrá siempre un grado de incertidumbre estadística. La intensidad de una fuente radiactiva viene dada por el número de partículas emitidas por unidad de tiempo. En cualquier caso, sólo seremos capaces de computar medias globales.

1.2 INTERACCION DE LOS RAYOS X Y GAMMA CON LA MATERIA

La detección de la radiación es posible debido a que la interacción de la radiación con la materia produce fotones, pares ión-electrón o pares electrón-hueco. El número de iones, fotones, o pares electrón hueco depende de la fracción de energía de la radiación consumida en el volumen sensible, de las propiedades del material y, algunas veces, de la naturaleza de la radiación. En el diseño de detectores y en el análisis de la información obtenida de ellos, es importante conocer la rapidez con la que la radiación cede su energía en el medio, la relación entre el número de pares y la energía ..., en fin, la naturaleza de la interacción para saber interpretar el espectro de la emisión. Discutimos ahora estos aspectos referidos sobre todo a la radiación gamma y los rayos X.

Los rayos X y los rayos gamma son radiación electromagnética. Considerados como partículas, podemos considerar que tienen masa y carga igual a cero. Un nombre común para los rayos X y gamma como partícula es el de *fotón*. La relación entre su energía y frecuencia viene dada por la conocida relación:

$$E = h\nu = h\frac{c}{\lambda} \quad (1.1)$$

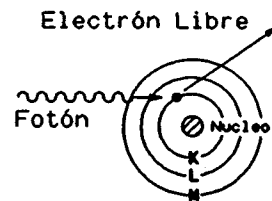
No hay una frontera clara entre ambas radiaciones. El término "rayos X" se aplica a los fotones con $E < 1 \text{ MeV}$. Gamma son los fotones con $E > 1 \text{ MeV}$. Hay una larga lista de posibles interacciones de fotones, pero las más importantes van a ser discutidas ahora: el efecto fotoeléctrico, el efecto Compton y la producción de pares ([12]).

El *efecto fotoeléctrico* es una interacción entre un fotón y un electrón. Como resultado del choque, el fotón desaparece, comunicando toda su energía al electrón, que es expulsado del átomo como un electrón libre o *fotoelectrón* (Fig. 1.1a). La energía cinética del electrón es $T = E_\gamma - B$, donde E_γ es la energía del fotón y B la energía de enlace del electrón.

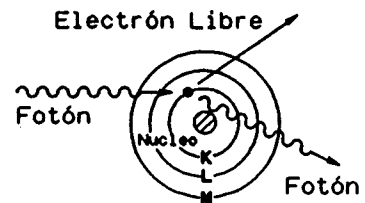
El *efecto Compton* es una colisión entre un fotón y un electrón libre. Por supuesto, en circunstancias normales, los electrones de un medio no están libres, sino ligados. Sin embargo, si la energía del fotón es del orden del KeV., el electrón puede ser considerado libre. El fotón no desaparece tras un efecto Compton. Sólo su energía y su dirección de movimiento cambian (ver Fig. 1.1b). La energía del fotón disminuye en una cantidad que se le entrega al electrón.

La *producción de pares* es una interacción entre un fotón de energía mayor a 1.02 MeV. y un núcleo. Como consecuencia de la interacción, el fotón desaparece y aparece un par electrón-positrón (Fig. 1.1c). Para energías mayores de 2 MeV, este es el proceso de interacción más importante.

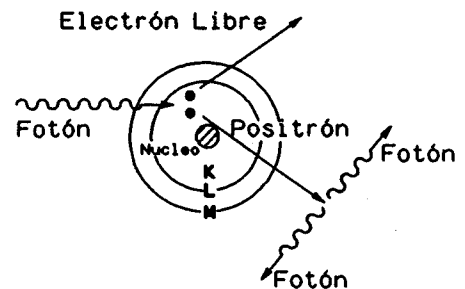
Cuando un fotón viaja a través de la materia, puede interaccionar de cualquiera de las tres formas antes discutidas. La Fig. 1.2 ilustra la importancia relativa de las tres interacciones en función de E_γ y Z , la energía del fotón y el número atómico del material que lo absorbe ([13]). Consideremos un fotón con $E = 0.1$ MeV. Si esta partícula incide en el carbono ($Z = 6$), el efecto Compton es el mecanismo de interacción predominante. Si el mismo fotón interacciona con yodo ($Z = 53$), el efecto fotoeléctrico prevalece. Para una partícula gamma de 1 MeV, el efecto Compton predomina con independencia de Z . Si un fotón de 10 MeV interacciona con el carbono, el efecto Compton será el fenómeno casi exclusivo. El mismo fotón en yodo será interceptado, principalmente mediante producción de pares.



a) **Efecto Fotoeléctrico**



b) **Efecto Compton**



c) **Creación de pares**

Fig. 1.1 Modos de interacción de la radiación X y gamma con la materia ([12]).

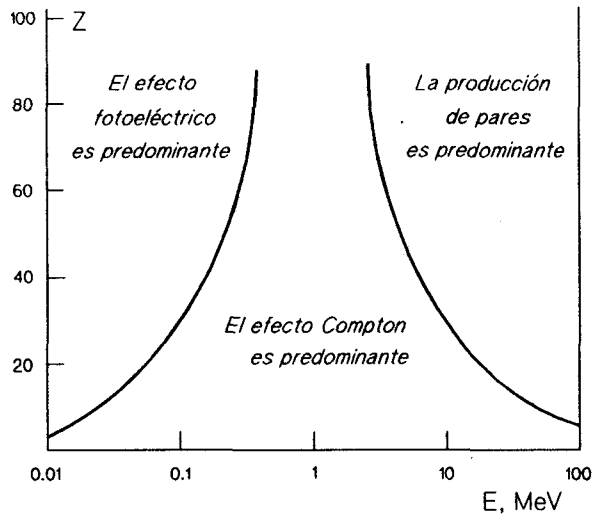


Fig. 1.2 La importancia relativa de las tres principales interacciones gamma ([13]).

La probabilidad total de interacción es la probabilidad de que se produzca una interacción de cualquiera de las tres modalidades por unidad de distancia. Si una emisión monoenergética de rayos gamma de intensidad I_0 atraviesa un material de espesor T , la fracción que atraviesa el medio sin ninguna interacción es igual a $e^{-\mu T}$ (Fig. 1.3), siendo μ el llamado el *coeficiente de atenuación lineal* del material. La probabilidad de que un fotón no interactúe con un material de espesor T es:

$$\frac{n^\circ \text{ de partículas transmitidas}}{n^\circ \text{ de partículas incidentes}} = \frac{I_0 e^{-\mu T}}{I_0} = e^{-\mu T}$$

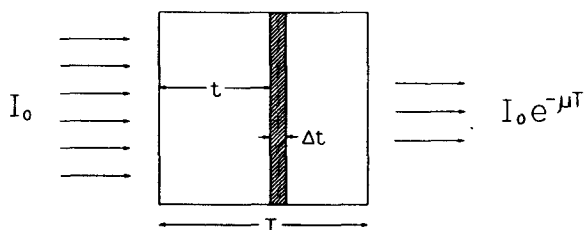


Fig. 1.3 La intensidad de la radiación transmitida (sólo los fotones que no interactuaron) decrece exponencialmente con el espesor del material.

1.3 INTRODUCCION A LOS DETECTORES DE RADIACION

La historia de los detectores de radiación nuclear corre paralela con el crecimiento de la física atómica y subatómica. Un ejemplo es el descubrimiento por Röntgen de los rayos X en 1895 usando emulsiones fotográficas. Desde entonces, las emulsiones han sido ampliamente empleadas como detectores de radiación. Al año siguiente, Bequerel descubrió la radiactividad con la ayuda de placas fotográficas. Rutherford utilizó una pantalla fosforescente y un microscopio para ver los destellos de luz producidos por partículas alfa en sus famosos experimentos. Su descubrimiento del núcleo atómico estuvo basado en ellos. Detectores similares conocidos como espintaroscopios, desarrollados por Crooks en 1903 fueron los precursores de los detectores de centelleo.

También se descubrió que las radiaciones producían ionización en el aire. Electroscopios y electrómetros fueron usados en el descubrimiento de los rayos cósmicos. Marie Curie usó electrómetros en sus estudios de la radiactividad de una amplia variedad de sustancias. Usando estos simples instrumentos, llegó a la conclusión de que la actividad era proporcional a la cantidad de sustancia radiactiva presente.

La extrema dificultad de contar chispas fosforescentes animó a Rutherford y a

Tabla 1.1

NOMBRE	TIPO	INTERACCION	MEDIO
Contador de ionización	Señal	Ionización	Gas
Contador proporcional			
Contador Geiger			
Contador de centelleo		Producción de fotones	Gas, líquido y sólido
Contador de semicon- ductor		Producción de pares electrón hueco	Sólido
Contador de Cherenkov		Producción de fotones por efecto Che- renkov	Gas, líquido y sólido
Emulsión fotográfica	Trazas	Ionización	Sólido
Cámara de nube			Gas
Cámara de burbujas			Líquido
Cámara de chispas			Gas y sólido
Detectores fotocrómicos		Cambio en la oxidación de iones de hierro	Sólido

Geiger a buscar un detector mejor. Dos hechos fueron fundamentales. Primero, las radiaciones producían iones. Segundo, los iones se multiplican en la presencia de un campo eléctrico suficientemente alto. En 1908 anunciaron un detector operativo. Desde entonces se han producido varios avances importantes en los métodos de detección de la radiación nuclear ([5], [8], [13]).

La Tabla 1.1 muestra los diferentes tipos de detectores usados en la actualidad. Los contadores de traza son de interés para los físicos de partículas de altas energías. Los contadores de tipo señal son de interés para radioquímicos, físicos de partículas de baja energía, biólogos, bioquímicos y geólogos que utilizan técnicas de radioisótopos.

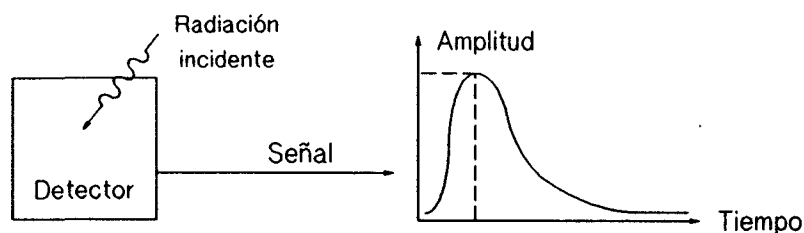


Fig. 1.4 Principio de operación de un contador de pulsos. Variación de la amplitud de pulso como una función del tiempo que sigue a la incidencia de una partícula en el detector.

1.3.1 CONCEPTOS BASICOS

La operación de un contador de pulsos típico se indica en la Fig. 1.4. Una partícula cargada incide sobre el detector y origina una señal o pulso a la salida. Las señales llegan de forma aleatoria en el tiempo, pero la frecuencia media de ocurrencias está determinada por la intensidad de la radiación. La proporcionalidad entre la amplitud de la señal o altura del pulso y la energía de la radiación, la eficiencia del contador, la resolución de energía y la anchura del pulso son algunos de los factores que se consideran al escoger un tipo u otro de detector. Discutimos ahora las características que influyen en la elección de un contador.

La **eficiencia** puede ser definida como la razón del número de señales detectables producidas y el número de las radiaciones incidentes. Algunas veces, especialmente para radiaciones de baja energía, la amplitud de las señales producidas puede ser más pequeña que el ruido, siendo así imposible detectarlas. Un factor influyente en este sentido es la electrónica empleada para analizar las señales. Esta tiene límites de sensibilidad, por lo que pulsos con amplitudes por debajo del límite inferior no serán detectados. Por lo tanto, la eficiencia de un contador variará con los límites inferiores de la electrónica que sigue al detector.

La **resolución de energía** de un contador es la medida de su capacidad para dis-

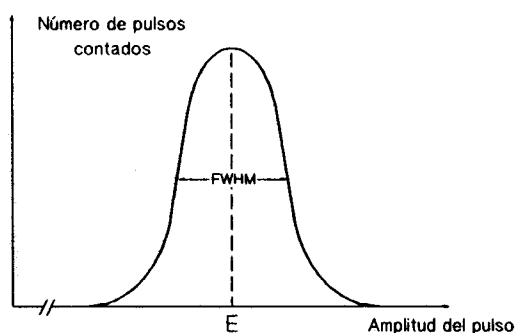


Fig. 1.5 Distribución típica de acumulación de pulsos de una radiación monoenergética E . El parámetro $FWHM$ determina la resolución del detector

criminar entre radiaciones de energías cercanas, por lo que es el parámetro más importante de un espectrómetro. En un sistema ideal, la absorción de fotones de la misma *energía* produce pulsos de la misma *amplitud* y la electrónica alimenta el analizador multicanal con señales idénticas, lo que provoca una acumulación de cuentas en un *canal* único correspondiente a esa amplitud de pulso. Los espectros resultantes de este sistema ideal son espectros de rayas de la anchura de un sólo canal. En la práctica, las líneas espectrales se dispersan a lo largo de varios canales como consecuencia de que los procesos de recolección de carga en el detector siguen una pauta estadística y a que, además, posteriormente el ruido de la electrónica de tratamiento del pulso contribuye a esta dispersión en canales. Una distribución típica de altura de pulso se muestra en la Fig. 1.5.

La duración del pulso en el tiempo es otro de los factores importantes que usualmente son considerados al seleccionar un detector. Si la duración es grande se provoca el solapamiento de pulsos o efecto de *apilamiento*, con el consiguiente error en la asignación de canal al pulso. El pulso tiene dos partes: la rampa de subida y la pendiente de bajada. La primera está determinada por las propiedades del contador (tales como el tiempo de recolección de iones o pares electrón-hueco) y la segunda depende de la constante RC del sistema. La anchura total puede ser reducida disminuyendo la constante de tiempo; sin embargo, una relación apropiada entre el tiempo de recolección y el tiempo de bajada debe ser mantenida para conservar la proporcionalidad entre la energía y la altura del pulso.

La amplia gama de energías cubierta por los espectrómetros nucleares -desde rayos X de 1 KeV hasta partículas de 100 MeV- hace imposible diseñar un sistema óptimo para el rango completo ([1]). La incidencia de un parámetro dado cambia radicalmente con la energía. Por ejemplo, el ruido electrónico es el factor dominante en la resolución a bajas energías, pero, a altas energías, su efecto es despreciable con respecto a los fenómenos estadísticos en el detector. Estas consideraciones conducen a establecer, si bien con fronteras a veces poco precisas, tres grupos de espectrómetros ([1]):

- 1- Espectrómetros de partículas de alta energía. Se caracterizan por su sensibilidad a factores como atrapamiento de cargas en el detector y derivas en ganancia de amplificadores, ya que la importancia de estos efectos aumenta linealmente con la energía. El ruido de la electrónica es prácticamente despreciable. Se emplean detectores de silicio en estas aplicaciones a temperatura ambiente.
- 2- Espectrómetros de rayos gamma. Miden energías mayores 30 KeV y usan detectores de Ge o INa, ya que el silicio no es eficiente a estas energías. Todos los factores que juegan en contra de la resolución, ruido electrónico, estadística del detector y atrapamiento de cargas, tienen importancia.
- 3- Espectrómetros de rayos X. Energías inferiores a 30 KeV. La resolución está dominada por el ruido electrónico. Para disminuir su efecto se emplean detectores de silicio a 77 °K con el fin de reducir la corriente de deriva, así como una electrónica sofisticada de bajo ruido.

1.4 DETECTORES DE GAS

Los detectores de gas operan utilizando la ionización producida por la radiación cuando esta pasa a través de un gas. Estos contadores tienen dos electrodos entre los que se aplica un potencial. El espacio entre ambos electrodos se llena con gas (Fig. 1.6). La

radiación pierde parte de su energía cuando pasa a través del espacio entre los electrodos, generando *pares ión-electrón*. Su movimiento induce una corriente en los electrodos que puede ser medida (Fig. 1.6a). A través de electrónica apropiada, la carga producida por la radiación puede ser transformada en un pulso, en cuyo caso las partículas son contadas individualmente (Fig. 1.6b). El primer tipo de contador se llama *cámara de corriente* y el segundo *cámara de pulso*.

1.4.1 RELACION ENTRE TENSION APLICADA Y CARGA RECOGIDA

Supongamos que realizamos el siguiente experimento (Fig. 1.7). Una fuente radiactiva de intensidad constante se ubica a una distancia fija del contador. Una fuente de alta tensión aplicada al contador puede ser variada mediante un potenciómetro. Si la tensión aplicada al contador se incrementa de forma constante, la carga recogida por unidad de tiempo cambia según muestra la Fig. 1.8. La curva de la Fig. 1.8 está dividida en cinco regiones:

Región I. Cuando el campo eléctrico en el contador es débil, los electrones y los iones se mueven con velocidades relativamente bajas y su tasa de recombinación es considerable. Al incrementar V , esta tasa descende hasta llegar a cero en $V = V_1$. La región I se llama *región de recombinación*.

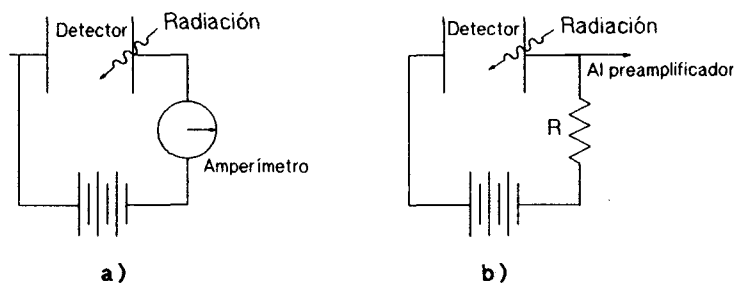


Fig. 1.6 Un detector de gas típico: a) se mide la corriente directa producida en el circuito. b) se detectan pulsos individuales ([13])

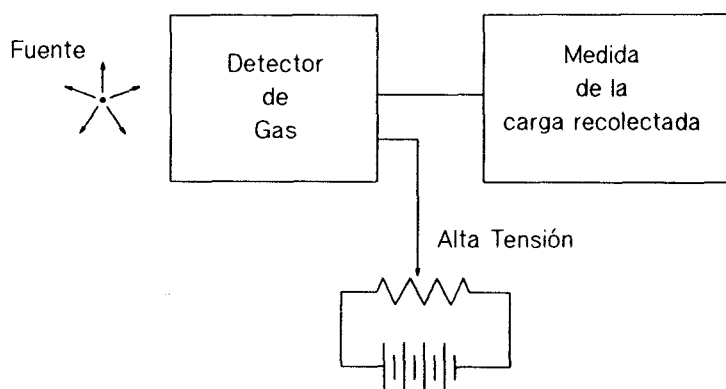


Fig. 1.7 Disposición experimental para el estudio de la relación entre tensión aplicada y carga recogida ([13]).

Región II. En esta región, la carga recogida permanece constante al subir la tensión porque la recombinación es ya nula y no se producen otras cargas. Se llama *región de ionización*.

Región III. El campo eléctrico es suficientemente fuerte para que los electrones producto de la ionización primaria adquieran energía suficiente entre colisiones para producir una ionización adicional llamada ionización secundaria. Para una tensión dada, la razón entre la ionización total y la ionización primaria es independiente de esta última. Esta razón se llama *factor de multiplicación del gas*. La salida del contador es, así, proporcional a la ionización primaria. La amplitud del pulso es proporcional a la energía disipada en el contador haciendo posible la medida de energía y la identificación de la partícula. Esta región se llama por esta razón *región proporcional*.

Región IV. En esta región el campo eléctrico es tan fuerte que un único par ión-electrón es suficiente como para provocar una avalancha de pares. Esta avalancha produce una señal muy fuerte con forma y amplitud independientes de la ionización primaria y el tipo de partícula y que únicamente depende de la electrónica del contador. Esta región se llama la región de *Geiger-Müller*.

Región V. Si la tensión aplicada va más allá de V_{IV} , un simple evento de

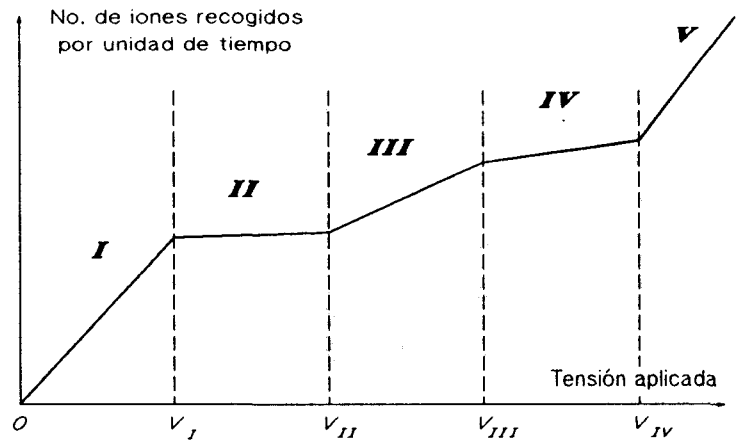


Fig. 1.8 Relación entre la tensión aplicada al contador y la carga recogida.

ionización inicia una descarga continua en el gas y el dispositivo deja de comportarse como un contador. Es más, un contador comercial operando en esta región seguramente se destruirá.

1.4.2 TIPOS DE CONTADORES DE GAS

Los contadores de gas toman su nombre de la región de tensión de la Fig. 1.8 en la que operan. Ningún contador opera en la región I porque un ligero incremento en tensión provocará un cambio en la señal.

Las **cámaras de ionización** operan en la región II. No tiene lugar ninguna multiplicación de carga. Es posible la medida de la energía de la partícula porque la señal de salida es proporcional a la energía depositada en el detector. Ya que esta señal no es grande, sólo es posible medir partículas fuertemente ionizantes tales como alfas, protones e iones pesados. La tensión aplicada es menor de 1000 V.

Los **contadores proporcionales** operan en la región III. Tiene lugar la multiplicación de la carga, pero la salida es aún proporcional a la energía depositada en el contador

y es posible la medida de la energía de la partícula. El contador proporcional se usa para detectar cualquier partícula cargada. La tensión aplicada oscila entre los 800 y los 2000 V.

Los **contadores de Geiger-Müller** operan en la región IV. Son muy útiles porque su operación es sencilla y la señal es tan fuerte que no es necesario el uso del preamplificador. Son sensibles a cualquier tipo de radiación ionizante. Su desventaja estriba en que la señal es independiente de la partícula y de su energía, por lo que proporciona información solamente sobre el número de partículas o actividad de la radiación. La tensión varía entre 1000 y 3000 voltios.

1.5 DETECTORES DE CENTELLEO

Los detectores de centelleo son materiales -líquidos, sólidos o gaseosos- que producen destellos o centelleos de luz cuando las radiaciones ionizantes pasan a través de ellos ([4], [12], [13]). El primer material sólido que se usó como detector de partículas fue un detector de centelleo. Fue usado por Rutherford en 1910 en su experimento sobre desviación de partículas alfa. Las partículas eran contadas a ojo con la ayuda de un microscopio, un proceso ineficiente e inexacto. El método se abandonó y fue resucitado 30 años después cuando avances en la electrónica hicieron posible la amplificación de la luz producida en el detector.

Cuando una partícula nuclear cargada, o un fotón de radiación X o gamma atraviesa ciertas sustancias fluorescentes, pierde energía en excitar su luminescencia dando lugar a la emisión de luz visible o ultravioleta. Dicha luz puede alcanzar el cátodo de una fotocélula suficientemente sensible, de la que desprende electrones, resultando un impulso eléctrico en respuesta a la llegada de la partícula o fotón. Se obtiene así un llamado *detector de centelleo* que consta por consiguiente de dos órganos fundamentales: la *sustancia luminiscente* y la *fotocélula*. El conjunto se monta en una cámara oscura en la que penetra la radiación que se desea detectar.

Las sustancias fluorescentes comúnmente empleadas en detectores de centelleo suelen ser o cristales inorgánicos o compuestos orgánicos, bien en forma cristalina o bien en disolución. Unos y otros deben ser transparentes a la luz que en ellos se genera por efecto de la radiación. Se les suele añadir en pequeña proporción una llamada *sustancia activadora* que aumenta la eficiencia de la detección.

La fotocélula u órgano encargado de convertir la luz generada en un impulso eléctrico es en realidad un llamado *fotomultiplicador* de extraordinaria sensibilidad. En su interior, en el que se ha hecho un vacío elevado, los electrones liberados por la luz en el fotocátodo, se multiplican al chocar sucesivamente con diversos electrodos metálicos llamados *dinodos* conectados a potenciales eléctricos positivos cada vez mayores. En cada uno de estos choques un electrón libera varios nuevos electrones, formándose una avalancha creciente de cargas eléctricas y, por tanto, un impulso más intenso de corriente. Se consiguen multiplicaciones electrónicas comprendidas entre 10^5 a 10^7 dependiendo del número de dinodos del tubo (entre 10 y 14) y de la diferencia de potencial entre cada par de dinodos consecutivos. La Fig. 1.9 representa esquemáticamente uno de estos detectores. Las sustancias fluorescentes inorgánicas más utilizadas en detectores de centelleo son el sulfuro de zinc activado con plata, $\text{SZn}(\text{Ag})$, y el yoduro de sodio activado con talio $\text{INa}(\text{Tl})$. Este último constituye el centelleador inorgánico más usado por su alta eficiencia

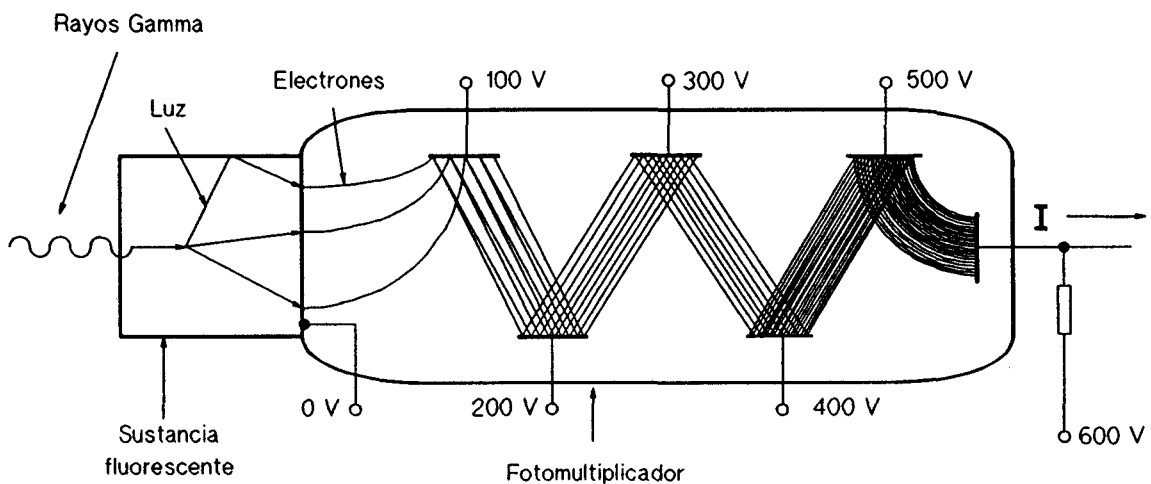


Fig. 1.9 Detector de centelleo ([12]).

de conversión luminosa. Se obtienen monocristales de INa(Tl) de tamaños muy variados, desde unos pocos centímetros cúbicos hasta varios decímetros cúbicos, con los que se consiguen eficiencias de detección gamma mucho mayores de las obtenidas con detectores de ionización gaseosa. Estos cristales, por ser muy higroscópicos y perder su transparencia con la humedad, deben montarse en cápsulas herméticamente cerradas, con una de sus paredes de vidrio transparente para dar salida a la luz hacia el cátodo del tubo fotomultiplicador.

Cuando un fotón de radiación X o gamma es enteramente absorbido en el cristal de centelleo, el impulso eléctrico resultante en el fotomultiplicador resulta ser de amplitud proporcional a la energía del fotón absorbido. Ello hace posible el uso de estos detectores para obtener el espectro de un haz de radiación gamma y deducir la naturaleza del radioisótopo emisor. De esta cuestión nos ocuparemos en los siguientes capítulos.

Diversas sustancias orgánicas fluorescentes como el antraceno, estilbeno, terfenilo, xileno, etc., son empleadas en detectores de centelleo, frecuentemente mezclados, y en forma líquida tras su disolución en disolventes apropiados. Estos centelleadores, así dispuestos, son particularmente apropiados para la medida de emisores de radiación beta. Se usan ampliamente para medir actividades débiles de emisión como las del carbono-14 y tritio, radioisótopos muy empleados en aplicaciones químicas y biológicas.

En los detectores de ionización gaseosa vistos anteriormente, los iones y electrones liberados en el proceso de detección deben moverse hacia los electrodos chocando con los átomos o moléculas del gas. Dicho movimiento no puede ser muy rápido, sobre todo para los iones por tener una masa mucho mayor. Por esta razón, el tiempo muerto de los detectores de centelleo es notablemente menor que el de los de ionización gaseosa, pudiendo detectar separadamente partículas o fotones que lleguen separados por un tiempo muy corto, de sólo una fracción de microsegundo. Así, con un detector de centelleo se puede contar, sin error apreciable, hasta decenas de miles de partículas o fotones por segundo.

1.6 DETECTORES DE SEMICONDUCTOR

La operación de un detector semiconductor es análoga a la operación de una cámara de ionización. En una cámara de ionización, la radiación incidente produce iones positivos y electrones y al recoger estos iones se produce un pulso eléctrico. En un contador de semiconductor, la radiación incidente produce pares electrón-hueco. Una diferencia fundamental entre ambos es que en un semiconductor se consume una energía de alrededor de 3.5 eV para provocar un par electrón-hueco, mientras que este valor se eleva a 30 eV en una cámara de ionización para crear un par ión-electrón. Este notable ahorro incrementa el número de pares por MeV de radiación y aumenta, como después veremos, su *resolución de energía* o habilidad para resolver la energía de las partículas de una fuente de emisión polienergética. Otras ventajas son la *respuesta lineal*, la *mayor eficiencia*, su rampa de subida de pulso rápida, la posibilidad de operación en el vacío y su insensibilidad a los campos magnéticos.

Las características de un detector de semiconductor no dependen solamente del tipo de material utilizado, sino también de la forma de crecimiento del cristal y de su tratamiento.

1.6.1 OPERACION DE UN DETECTOR SEMICONDUCTOR

Para comprender la operación de un contador de semiconductor, consideremos una barra de un cristal uniforme con electrodos incorporados en ambos extremos según la Fig. 1.10. Cuando la radiación incide en el detector se producen pares electrón-hueco y cuando éstos son recogidos en los electrodos se obtiene una señal en la resistencia cuya amplitud es *proporcional* a la energía de la radiación cedida al material. El número de electrones producidos por la radiación debería ser mucho mayor que el número de electrones que están ya en la banda de conducción o, en otras palabras, el nivel de continua de la corriente debiera ser muy pequeño. Esto asegurará que las variaciones aleatorias de la corriente no

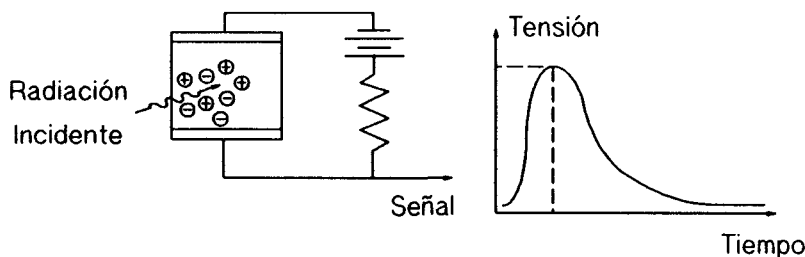


Fig. 1.10 Contador de semiconductor de un sólo cristal.

sean confundidas con el pulso inducido por radiación. Por lo tanto, un buen material sensor sería aquél que tuviese una baja conductividad, preferiblemente un aislante perfecto a la temperatura de operación. Este requisito elimina los metales. Nos quedan los semiconductores y los aislantes.

La resolución en energía es generalmente un factor importante al seleccionar un material detector. La resolución depende del número de pares electrón-hueco producidos por la radiación. La energía necesaria para producir un par electrón-hueco es más baja en un semiconductor que en un aislante. Por lo tanto la resolución obtenida en un semiconductor será mucho mayor que la obtenida en un aislante. Si queremos resolución de energía, sólo nos quedan los materiales semiconductores.

Los pares electrón-hueco producidos deberían moverse libremente en el detector para alcanzar los electrodos. Por lo tanto, es conveniente usar cristales con el mínimo número de trampas posible. Los centros de atrapamiento mejoran la recombinación de huecos y electrones, por lo que también disminuirán la amplitud del pulso. Para discutir todos estos factores examinemos los diagramas de niveles de energía del semiconductor y del aislante de la Fig. 1.11. La anchura de la banda prohibida es 0.69 y 1.08 eV para el germanio y el silicio respectivamente. Debido a la energía térmica, algunos electrones son capaces de saltar a la banda de conducción. La probabilidad de que un electrón salte a la

banda de conducción es proporcional al factor $e^{-\frac{E_g}{2kT}}$, donde E_g es el ancho de energía de la banda prohibida del material. De acuerdo con la forma exponencial de este término, siempre hay electrones en la banda de conducción. Estos electrones son los responsables

de la corriente de deriva o del indeseable nivel de continua antes aludido. Esta corriente disminuye al hacerlo la temperatura. A temperatura ambiente ($T = 300\text{ }^{\circ}\text{K}$), la densidad de la nube electrónica de la banda de conducción para el silicio es de 10^{10} electrones por cm^3 y para el germanio vale 2×10^{13} . La desviación típica de esta distribución es para el silicio $(10^{10})^{1/2} = 10^5$, que es el mismo número de electrones producidos por 1 MeV de radiación. Por lo tanto, es difícil discriminar entre los pulsos producidos por la radiación y los debidos a las fluctuaciones estadísticas de la concentración de la nube electrónica de la banda de conducción. De esta discusión es fácil deducir que una banda prohibida ancha es útil para reducir pulsos espúreos debidos al ruido térmico pero inapropiada para alcanzar una buena resolución energética. Podemos asimismo deducir que un buen detector exige que no existan más pares electrón-hueco libres que los que forman las radiaciones que tratamos de detectar. Esto se consigue empleando un material muy puro y bajando la temperatura. Dada la dificultad práctica que presentan estos requisitos, se recurre a la técnica empleada en la fabricación de diodos y transistores a fin de obtener zonas de semiconductor sin cargas eléctricas libres a temperaturas normales. Así, en la unión de un semiconductor ligeramente contaminado con impurezas trivalentes (tipo *P*) con el mismo semiconductor contaminado con impurezas pentavalentes (tipo *N*) se forma una zona libre de cargas y apta para la detección obteniéndose un llamado detector de unión.

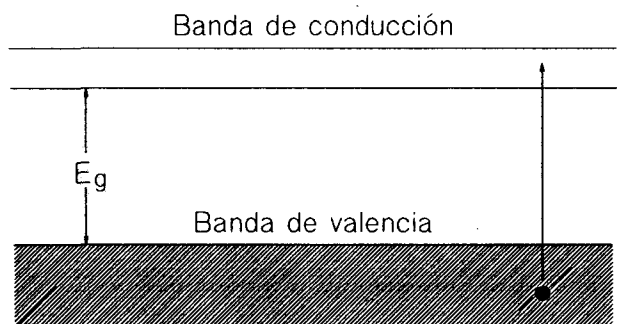


Fig. 1.11 Diagrama de niveles de energía de electrones en un semiconductor.

1.6.2 OPERACION DE LA UNION P-N COMO DETECTOR

Un semiconductor tipo N tiene un exceso de portadores de electrones. Uno de tipo P tiene exceso de huecos. Si ponemos en contacto un semiconductor de tipo P y otro de tipo N , los electrones y los huecos se mueven por dos razones:

- 1 Tanto los electrones como los huecos se moverán por difusión desde áreas de alta concentración a áreas de baja concentración.
- 2 Bajo la influencia de un campo eléctrico tanto los electrones como los huecos se moverán, pero en direcciones opuestas.

Consideremos dos semiconductores, uno de tipo P y otro de tipo N , en contacto sin la acción de un campo eléctrico externo (Fig. 1.12). La difusión se produce y, como resultado, la región de tipo N se cargará positivamente y la de tipo P negativamente. Un potencial se establece entre las dos regiones. Hemos formado una unión $P-N$.

El potencial V_0 (Fig. 1.12) depende de las concentraciones de electrones y huecos y es del orden de 0.5 V. Si se aplica una tensión externa V_b con el polo positivo en el lado

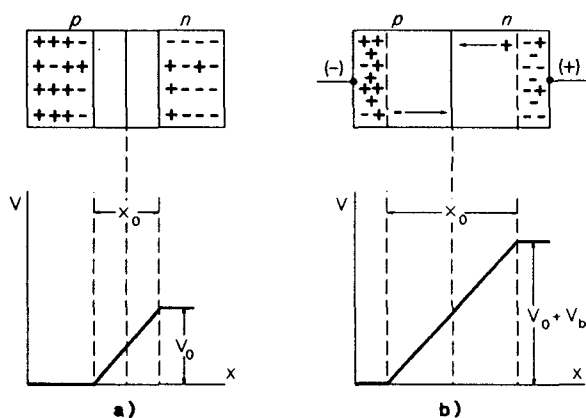


Fig. 1.12 (a) Una unión P-N sin tensión externa. (b) Aplicada una tensión inversa, el potencial en la unión aumenta así como la amplitud x_0 de la región de vaciamiento de portadores.

N , el potencial total a lo largo de la unión es de $V_0 + V_b$. Decimos que la unión está *polarizada en inversa*. Esta tensión externa tiende a hacer más difícil el movimiento de electrones y huecos. En la región de vaciamiento existe un campo eléctrico $E = -\partial V / \partial x$. La longitud $x_0 = x_n + x_p$ de esta región se incrementa con la polarización inversa siendo $x_p \approx (\mu_p \rho (V_0 + V_b))^{1/2}$ en la región tipo P y $x_n = (\mu_n \rho (V_0 + V_b))^{1/2}$ en la región tipo N donde ρ es la resistividad del cristal, μ_n la movilidad de los electrones y μ_p la movilidad de los huecos. La aplicación de un potencial negativo en el lado N tendrá el efecto contrario. Esto se llama polarización directa. La diferencia de potencial total será de $V_0 - V_b$. Los detectores de radiación funcionan en polarización inversa. Ya que, generalmente, $V_b \gg V_0$, $x_0 \approx \sqrt{V_b}$ ((12)).

La operación de un detector semiconductor está basada en las propiedades de la unión $P-N$ en inversa (Fig. 1.13). La radiación incidente en la unión genera pares electrón-hueco cuando pasa a través de ella. Los huecos y los electrones son rápidamente desplazados por la influencia del campo eléctrico y con una electrónica apropiada la carga recogida produce un pulso que puede ser registrado. Ya hemos visto la necesidad de que en el detector no existan más huecos y electrones libres que los que generan las radiaciones que tratamos de detectar, por lo que el material semiconductor empleado debe ser de altísima pureza y operar a bajísima temperatura. Dada la dificultad práctica de estos requisitos, se recurre a la técnica empleada en la fabricación de diodos y transistores a fin de obtener zonas de semiconductor sin cargas eléctricas libres a temperatura ambiente. Así, en la unión de un semiconductor ligeramente contaminado con impurezas aceptadoras (semiconductor tipo P), y el mismo semiconductor contaminado con impurezas donadoras (tipo N), se forma una zona libre de cargas y apta para la detección llamada zona de vaciamiento o región sensible.

La bondad de un detector de este tipo depende de la región x_0 de la Fig. 1.12. Los electrones y huecos producidos en esa región se encuentran en un entorno similar al de los iones y electrones en un cámara de ionización. Hay, sin embargo, algunas diferencias entre ambos tipos de detector. En un contador de gas, la movilidad de los electrones es miles de veces más grande que la de los iones. En los semiconductores, la movilidad de los

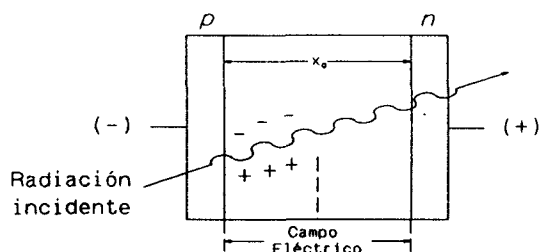


Fig. 1.13 La unión p-n con polarización inversa operando como detector de radiación.

electrones es dos o tres veces mayor que la de los huecos. El tiempo empleado en recoger la carga en un contador de gas es del orden de milisegundos. En los semiconductores, la región sensible del contador tiene solamente unas cuantas micras y la velocidad de electrones y huecos es tal que los portadores de carga pueden atravesar la región sensible y ser recolectados en un tiempo del orden de 10^{-7} segundos.

El objetivo en contadores tanto de ionización como de semiconductor es recoger todas las cargas producidas por una partícula incidente. Esto se alcanza estableciendo un campo eléctrico en el detector tal que la recombinación de iones o huecos sea despreciable antes de ser recogidos. En un detector de semiconductor, incluso si la recombinación es cero, algunos portadores de carga pueden "perderse" en centros de atrapamiento del cristal tales como imperfecciones en la red, vacantes y dislocamientos. Además, la radiación incidente crea defectos en el cristal que causan un deterioro en la calidad del mismo y reduce así su tiempo de vida útil.

La capacidad de la unión P-N es importante porque afecta a la resolución de energía del detector. La capacidad de transición C de un detector viene dada por:

$$C = \epsilon \frac{A}{4\pi x_0}$$

donde:

ϵ = constante dieléctrica del material.

A = Area de la superficie del detector.

x_0 = Anchura de la región de depleción.

Según el valor de x_0 dado anteriormente puede demostrarse que $C \sim \frac{1}{\sqrt{V_b}}$, donde

V_b representa la tensión externa aplicada.

Para resumir, un material adecuado para la construcción de un detector debe presentar ciertas propiedades. Las más importantes son las siguientes:

- 1 **Alta resistividad.** Es esencial, ya que de otra manera circulará corriente bajo la influencia del campo eléctrico y la carga producida por la radiación producirá un pulso que será fácilmente enmascarado por la corriente continua.
- 2 **Alta movilidad de portadores.** Los electrones y huecos deben ser capaces de moverse rápidamente y ser recogidos antes de recombinarse o ser atrapados. Este requisito entra en conflicto con el anterior porque en los materiales de alta resistividad la movilidad es baja. Los materiales semiconductores contaminados con impurezas han resultado ser el mejor compromiso.
- 3 **Capacidad para soportar fuertes campos eléctricos.** Esta propiedad está relacionada con la primera. Su importancia radica en el hecho de que cuanto mayor sea el campo eléctrico mejor y más rápida es la recolección de cargas. A mayor campo eléctrico mayor amplitud de la región sensible.
- 4 **Red cristalina perfecta.** Aparte de las impurezas inyectadas externamente, la red del cristal debe ser lo más perfecta posible. Los defectos, de la naturaleza que sean, actúan siempre como centros de atrapamiento de portadores.

1.6.3 DISTINTOS TIPOS DE SEMICONDUCTOR

Los distintos tipos de detectores semiconductores que existen hoy día difieren entre sí en el material usado en su construcción o en el método con el que se trata el detector. Esta sección revisa el método de construcción y las características de los dos detectores más utilizados en la práctica y dos detectores aún objeto de investigación: los de telurio de cadmio (CdTe) y los de yoduro de mercurio (I_2Hg).

1.6.3.1 DETECTORES DE DIFUSION DE UNION

El silicio de tipo *P* es el material básico de esta clase de detectores. La pieza de material tipo *P* tiene la forma de una oblea fina. Se forma una capa muy fina de silicio de tipo *N* en la cara de arriba de la oblea aplicando un compuesto de fósforo a la superficie y calentando el conjunto hasta los 800° o 1000 °C durante menos de una hora. El fósforo se difunde en el silicio y lo contamina con donantes negativos. Tenemos una unión *P-N* (Fig. 1.14).

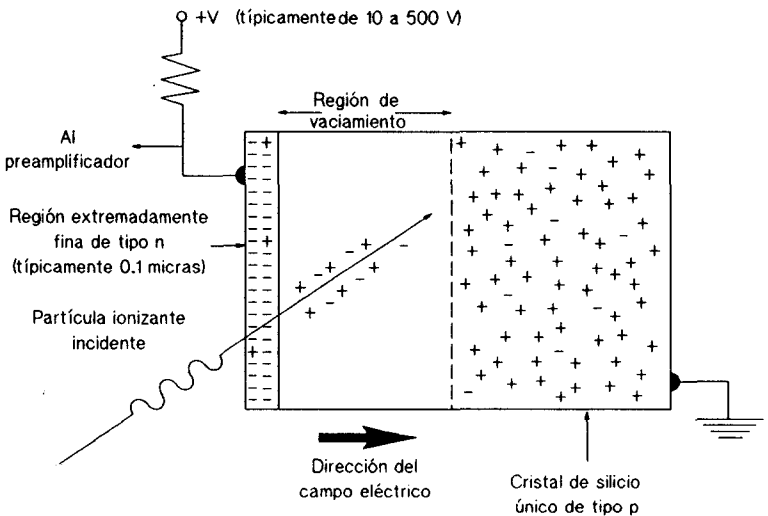


Fig. 1.14 Un detector de difusión de unión.

1.6.3.2 DETECTORES DE SILICIO COMPENSADOS CON LITIO

[Si(Li)]

Para el detector de difusión de unión, la región sensible, es decir, el tamaño real del detector tiene un límite superior de alrededor de 2000 micras. Esta limitación afecta a la energía máxima de la partícula que va a ser medida. La longitud de la región sensible puede aumentarse si difundimos iones de litio desde la superficie hasta el fondo de la oblea. Este proceso ha sido llevado a cabo con éxito en silicio y germanio generando los llamados Si(Li) y Ge(Li). La anchura del detector se ha elevado hasta 5 mm en el caso del silicio y hasta 12 mm en el caso del germanio.

El proceso de compensación con litio consiste en dos pasos principales. Uno, la formación de la unión $P-N$ mediante la difusión del litio. Dos, aumento de la región sensible mediante compensación de iones.

El método más simple de llevar a cabo la difusión consiste en depositar una suspensión de litio en aceite sobre la superficie de la oblea. Otros métodos son la deposición en el vacío o la electrodeposición. Después de que el litio es depositado en la superficie, la oblea se calienta durante 3 a 10 minutos a 250-400 °C en una atmósfera inerte como helio o argón. Cuando la difusión comienza, la concentración del receptor (N_p) es constante en todo el cristal (Fig. 1.15a), mientras que la concentración del donante (N_n) es alta en la superficie y nula en el fondo. A la profundidad x_j donde $N_n(x_j) = N_p$ se forma una unión $P-N$.

La compensación se lleva a cabo calentando la unión hasta 120-150 °C mientras se aplica polarización inversa entre 25 y 1000 V. El campo eléctrico tiende a llevar a los átomos tipo N (el litio) hacia el lado P de la unión y se crea una región intrínseca donde $N_p \approx N_n$. Para tiempos largos de compensación, el espesor de esta región $X_0(t)$ viene dado por $x_0 = \sqrt{2V\mu_L t}$ donde V es la tensión aplicada y μ_L la movilidad de los iones de litio a la temperatura de compensación. La compensación es un proceso largo. Dependiendo del grosor deseado puede llevar días y a veces semanas. Una vez completa la compensación,

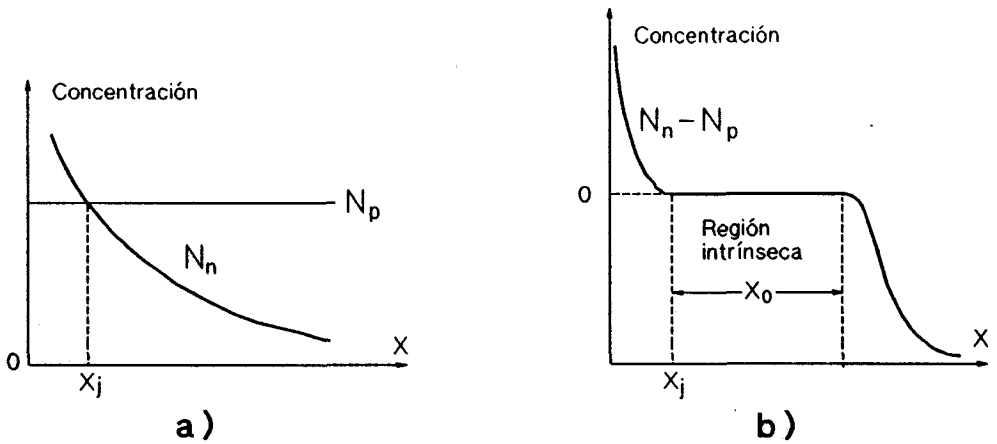


Fig. 1.15 (a) Difusión: la concentración del litio donante cambia con la profundidad. (b) Compensación: se forma una región casi intrínseca con amplitud x_0 .

el detector de Si(Li) se monta sobre un criostato porque los resultados mejores se obtienen cuando el detector opera a una temperatura muy baja. Generalmente la del nitrógeno líquido, que es de 77 °K. La baja temperatura es necesaria porque a temperatura ambiente la movilidad del litio es tal que su continua difusión arruinaría el detector. Los detectores de Si(Li) se usan para detección de partículas cargadas y especialmente rayos X.

1.6.3.3 DETECTORES DE GERMANIO COMPENSADOS CON LITIO [Ge(Li)]

Los detectores de germanio contaminado con litio [Ge(Li)] se forman mediante un crecimiento horizontal de cristales de germanio. Cuando el cristal está en proceso de crecimiento, se le contamina con impurezas aceptoras tales como el indio, el galio o el boro, y se transforma en un semiconductor tipo P . La compensación del germanio sigue la misma aproximación que la del silicio. Igualmente, la movilidad del litio en el germanio se incrementa con la temperatura. En cuanto a la velocidad de compensación, lleva cuatro semanas alcanzar una amplitud de 10 a 15 mm. Una vez terminado el proceso de compensación, el detector también se monta en un criostato a 77 °K. La conservación a esta temperatura es mucho más crítica que en el silicio. La movilidad de los átomos de litio en el germanio es tan alta a temperatura ambiente que el detector se deterioraría completamente

incluso en un lapso de tiempo muy corto. Debido a su estrecha banda prohibida (0.67 eV), aun en el caso de que existiere un mecanismo para alcanzar la estabilidad de compensación sería preciso operar a baja temperatura. Los cristales de Ge(Li) se usan para detectar radiación gamma.

1.6.3.4 EL YODURO DE MERCURIO Y EL TELURO DE CADMIO

La principal desventaja de los detectores compensados con litio es la necesidad de enfriamiento continuo. Requieren un criostato, que hace al contador muy voluminoso y por lo tanto imposible de usar en casos donde se dispone de un espacio limitado. Por supuesto, una limitación adicional es el suministro continuo de nitrógeno líquido. Existe por lo tanto un gran incentivo para desarrollar detectores de semiconductor que puedan ser almacenados y operados a temperatura ambiente. Dos materiales que están actualmente en estudio y se muestran prometedores para ser usados comercialmente como detectores son el CdTe y el I₂Hg ([2], [14]). Hasta 1991 han sido construidos detectores de 0,7 mm y área de 100 mm². Estos detectores son pequeños en tamaño comparados con los detectores de Ge(Li) y Si(Li), pero el volumen requerido depende del cometido. Estos cristales encuentran aplicación en monitorización de radiación en el espacio, medida de actividad en centrales nucleares y protección radiológica, sistemas portátiles de diagnóstico médico o sistemas de visión de radiaciones ([3], [7], [10], [11]). Debido al alto número atómico de los elementos implicados, su eficiencia es considerable a pesar de su reducido volumen. Ya que la energía precisa para producir un par electrón-hueco en estos detectores es mayor que en el silicio o el germanio, su resolución energética es menor. Ante este panorama, el uso de estos detectores se inclina hacia medidas donde la resolución energética no sea el requisito fundamental sino que se pueda establecer un buen compromiso entre resolución y, al mismo tiempo, pequeño volumen y operación a temperatura ambiente.

1.7 INSTRUMENTACION BASICA DE MEDIDA DE RADIACION

Vamos a presentar en este apartado una descripción general de los componentes básicos de un sistema de detección de radiación. Cada uno de ellos será tratado a modo de caja negra sin entrar en detalles internos de operación ([12]).

Los sistemas de detección que trataremos son de tipo pulso de tensión. Emiten un pulso por cada partícula detectada. La medida de radiación se encarna así como un mero mecanismo de cuenta de pulsos. La Fig. 1.16 muestra un sistema básico de conteo. La función de cada uno de los componentes se describe a continuación.

1.7.1 POLARIZACION DEL DETECTOR

El detector necesita ser polarizado mediante la aplicación de una tensión que depende de sus características (desde alrededor de 10 V en detectores muy finos hasta 3000 V en cristales contaminados con Litio). La fuente de tensión debe proporcionar una salida extremadamente estable con independencia de fluctuaciones en su alimentación.

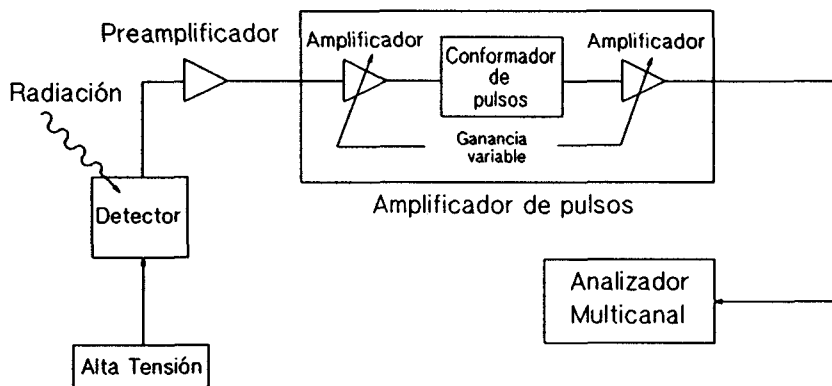


Fig. 1.16 Sistema básico de detección de partículas.

1.7.2 EL PREAMPLIFICADOR

La señal que sale del detector es del orden del milivoltio. Antes de registrarse debe ser amplificada por un factor del orden del millar o superior en el amplificador. Llevar el pulso hasta éste supone siempre el transporte de una señal débil por una línea de transmisión con los efectos de atenuación que ello comporta. El preamplificador lleva a cabo un acoplamiento de impedancias óptimo entre detector y amplificador de modo que la distorsión del pulso a la entrada de este último sea inapreciable. En la mayoría de los casos, la alta tensión se aplica al detector a través del preamplificador.

Un segundo propósito del preamplificador es minimizar fuentes de ruido que acompañan a la formación del pulso y que reducen la resolución global del sistema. Hay tres tipos de preamplificadores: sensibles a carga, a tensión y a corriente. El primero es el único utilizado en detectores de semiconductor. Se comprenderán mejor sus características observando la Fig. 1.17. En ella, tenemos un transistor de efecto de campo de bajo nivel de ruido de ganancia A que constituye la primera etapa amplificadora. El detector ve el transistor como un condensador de capacidad AC_f . Si $AC_f \gg C_i$, donde C_i es la capacidad del detector y del cable, la tensión a la salida del preamplificador será proporcional a la carga inducida en el detector por la radiación incidente:

$$V_0 = \frac{Q}{C_f}$$

1.7.3 EL AMPLIFICADOR

El amplificador tiene dos funciones. Una es el conformado de señal, que da una nueva forma al pulso a base de diferenciaciones e integraciones del mismo por medio de asociaciones RC. El conformado evita el solapamiento de pulsos, mejora la relación señal-ruido y facilita manipulaciones específicas de pulsos. Sin embargo, su función principal es amplificar la señal que llega del preamplificador. Esta se halla en el margen de los

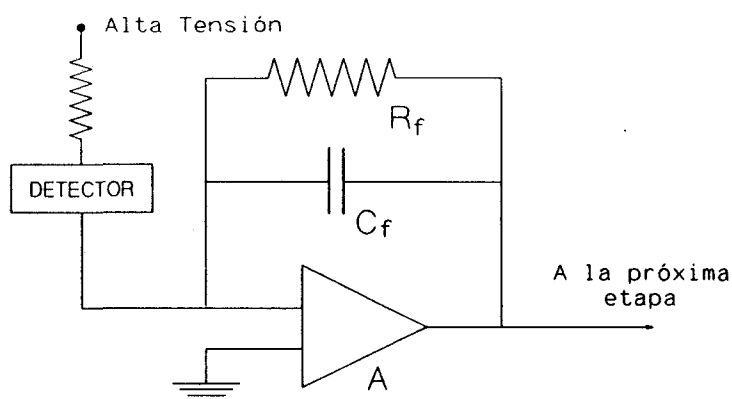


Fig. 1.17 Primera etapa de una preamplificador sensible a carga.

milivoltios y no puede llevarse muy lejos o ser manipulada sin que se pierda en el ruido de fondo. Los amplificadores modernos dan una salida entre en el margen [0-10] voltios. Esto significa que la ganancia debe ser cuidadosamente escogida a fin de mantener para toda la amplitud del pulso una función de transferencia lineal. El *control de ganancia* es su principal parámetro libre. Los amplificadores comerciales disponen de dos mandos de control para ajustar la amplificación. El primero es un selector que ajusta la amplificación por pasos. Cada posición es una fracción de la amplificación máxima. El segundo avanza de forma continua alrededor del valor fijado por el primero.

Un amplificador debe cumplir ciertos requisitos para permitir una buena medida. Dependiendo del problema a tratar, algunos son más apropiados que otros. Por ejemplo, en una aplicación donde sólo interesa la tasa de actividad de una emisión radiactiva, es decir, el número de partículas incidentes por unidad de tiempo y no su energía, la precisión y la estabilidad del amplificador no son muy importantes. Es en aplicaciones de espectroscopía donde estos requisitos se hacen necesarios. Un cambio en ganancia da lugar a una desviación funcional del espectro de energías que hay que tomar en cuenta a la hora de examinarlo. Por ello, una calibración de ganancia previa a una serie de medidas debe hacerse siempre, y más a menudo cuanto mayor sea su inestabilidad.

Un amplificador de espectroscopía ideal debería proporcionar una amplificación constante para toda amplitud de pulsos sin distorsionarlos. Desafortunadamente, siempre

aparece alguna distorsión en los pulsos debido al ruido electrónico, derivas en la ganancia como efecto de la temperatura, apilamientos de pulsos y las limitaciones en la linealidad del amplificador. Las desviaciones en la ganancia afectan principalmente a la espectroscopía posterior como ya hemos dicho y son debidas en gran medida a los pequeños cambios que la variación de temperatura produce en las características de resistencias, condensadores, transistores, etc. El valor de la deriva en ganancia es siempre especificado por el fabricante del instrumento y tiene unos valores del orden de hasta el 0.005 por ciento por grado centígrado.

A pesar de que existen amplificadores comerciales diseñados para cumplir requisitos específicos en espectroscopía u otras áreas, las derivas persisten, afectando directamente a la forma del pulso.

1.7.4 CONVERSORES ANALOGICO-DIGITALES

La columna vertebral de un analizador multicanal es el conversor analógico digital (ADC). Esta unidad asigna un código digital a cada amplitud de pulso. La exactitud de un ADC se expresa en términos de su alinealidad diferencial e integral. La alinealidad diferencial viene dada por la diferencia máxima de amplitud entre dos intervalos de amplitud de pulso, referida a la amplitud media del intervalo. Se sitúa en instrumentos comerciales entre el 0.5 y el 1 por ciento. La alinealidad integral es la máxima distancia entre la altura de pulso que corresponde idealmente a un código según una relación lineal y la que realmente se observa referida a la máxima altura observada. Los ADCs comerciales tienen una alinealidad integral en torno al 0.05 por ciento en el 95-98% del margen.

1.7.5 EL ANALIZADOR MULTICANAL

Medir u obtener un espectro de energía de una fuente de radiactividad significa registrar una realización de la distribución estadística de la energía de las partículas que emite la fuente. La Fig. 1.18 muestra uno de estos espectros. El analizador multicanal lleva

a cabo esta función. La operación consiste en clasificar los pulsos que llegan del amplificador de acuerdo a su amplitud y acumular el número de pulsos de la misma amplitud en una casilla o dirección de memoria correspondiente, que se llama el *número de canal*. Esta colección de datos se denomina el *espectro de energía* de la fuente de radiación.

Cuando un pulso llega al ADC, éste cuantifica su amplitud y genera un código binario proporcional a ella que corresponde a un número de canal. Cada canal no es más que un contador que va a incrementarse en la unidad cada vez que un pulso le es asignado. El número de canales o tamaño del ADC determina la resolución del analizador. Esta cifra en ADC's comerciales es de 256, 512, 1024, 2048, 4096 o 8192 canales ajustables a 256, 512, 1024, etc.

El primer canal de la región usada se llama el *canal cero* y mide la duración del conteo en segundos. Con este dato reconstruimos la actividad de la fuente dividiendo las cuentas acumuladas en todos los canales (suma integral del espectro) por el contenido del canal cero. Normalmente podremos seleccionar el uso de cierta porción del multicanal. Si usamos la memoria entera o la primera mitad o el primer cuarto el canal cero es la dirección cero. Si en una memoria de 2046 canales utilizamos el segundo cuarto, el canal

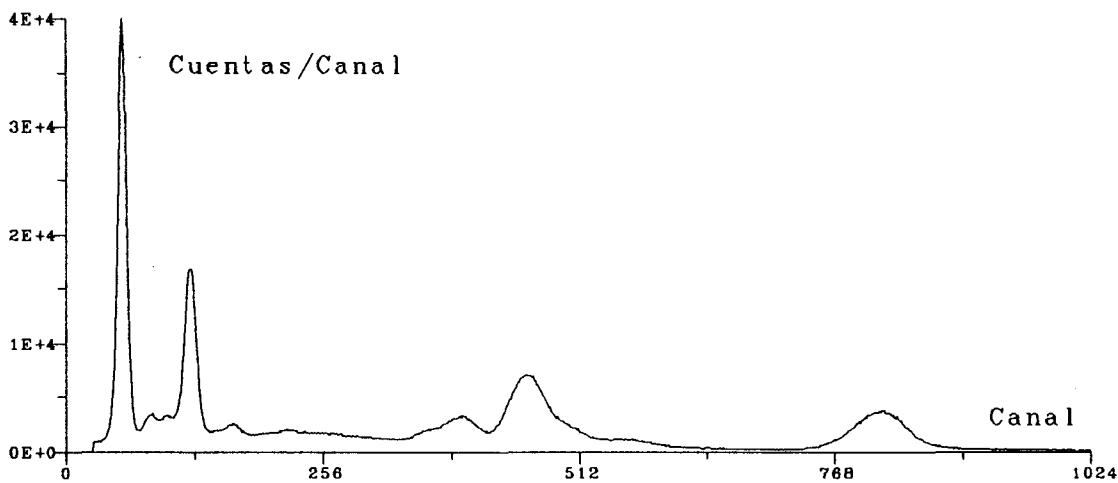


Fig. 1.18 Espectro resultado del conteo de una fuente emisora compuesta por ^{133}Ba y ^{137}Cs de un microcurio de actividad cada una. Se utilizó un espectrómetro de centelleo MCA Camberra 35 Plus.

cero es la posición 512 y así sucesivamente. Cuanto mayor sea la memoria utilizada mayor será la resolución alcanzada en el espectro.

1.7.6 CALIBRACION DE UN ANALIZADOR MULTICANAL

Cada número de canal se corresponde con un determinado intervalo de energía. Calibrar un analizador multicanal significa determinar en rigor su relación funcional. Esta ecuación puede aproximarse polinomialmente como

$$E = a_1 + a_2 C + a_3 C^2 + \dots \quad (1.2)$$

donde C es el número de canal y a_1, a_2, \dots son constantes.

Estas constantes son determinadas recogiendo espectros a partir de fuentes con picos de emisión de energía conocida. En principio, necesitamos, al menos, tantos picos como constantes para aplicar a la ecuación (1.2) el método de los mínimos cuadrados ([6]). Algunos autores ([9]) sugieren utilizar un único espectro con dos picos, uno de baja y otro de alta energía, ya que en la mayoría de los sistemas de detección, y sobre todo en los basados en cristal semiconductor ([6]), la ecuación (1.2) es esencialmente lineal:

$$E = a_1 + a_2 C \quad (1.3)$$

Las constantes a_1 y a_2 determinan los parámetros básicos de configuración del analizador multicanal. El primero suele ser pequeño y la mayoría de los instrumentos comerciales tienen un mando para ajustarlo a cero si así es requerido. El segundo viene dado por la ganancia global del preamplificador y el amplificador también modificable a voluntad. Ambas constantes sufren pequeñas desviaciones espontáneas a lo largo del tiempo debido a cambios en la temperatura y en las características de los diversos dispositivos del sistema global como ya hemos citado en el caso del amplificador.

1.8 BIBLIOGRAFIA

- [1] Cerny, J. (ed.), "Nuclear Spectroscopy and Reactions", Chapter III-A and III-D, Academic Press, (1974).
- [2] Entine, G. *et al.*, "Survey of CdTe Nuclear Detector Applications", Nuclear Instruments and Methods, A283, (1989), pp. 282-290.
- [3] Friant, A. *et al.*, "HgI₂ Detectors for Pu Contamination Monitoring and Radiological Protection Measurements", Nuclear Instruments and Methods, A283, (1989), pp. 227-231.
- [4] Heath, R. L. *et al.*, "Inorganic Scintillators. A Review of Techniques and Applications", Nuclear Instruments and Methods, 162, (1979), pp. 431-476.
- [5] Kandiah, K., "The evolution of Nuclear Detector Instrumentation", Nuclear Instruments and Methods, 162, (1979), pp. 699-718.
- [6] Mitchel, D. *et al.*, "Gamma-Ray Respose Functions for Scintillation and Semiconductor Detectors", Nuclear Instruments and Methods, A276, (1989), pp. 547-556.
- [7] Olmos, P. *et al.*, "Use of thick HgI₂ detectors as intelligent spectrometers", Nuclear Instruments and Methods for Physics Research, A299, (1990), pp. 45-50.
- [8] Ouseph, P.J., "Introduction to Nuclear Radiation Detectors", Plenum Press, (1975).
- [9] Parr, R. M. *et* Lucas, H. F., " A Rigourous Least Squares Analysis of Complex Gamma Ray Spectra with Partial Compensation for Instrumental Instability", IEEE Trans. on Nuclear Science, 1964, June, 349.
- [10] Patt, B. E. *et al.*, "Developments in Mercuric Iodide Gamma Ray Imaging", Nuclear Instruments and Methods, A283, (1989), pp. 215-219.
- [11] Piechotka, M. *et* Kaldis, E., "Mercuric Iodide Materials Research: Some Recent Developments and Open Problems", Nuclear Instrumets and Methods, A283, (1989), pp. 111-118.
- [12] Tanarro, A., "Radiaciones ionizantes", Publicaciones científicas del C.I.E.M.A.T., 1986.

- [13] Tsoulfanidis, N., "Measurement and Detection of Radiation", McGraw-Hill, 1983.
- [14] Whited, R. C. *et* Schieber, M. M., "Cadmium Telluride and Mercuric Iodide Gamma Radiation Detectors", Nuclear Instruments and Methods, 162, (1979), pp. 113-123.

Capítulo 2

EL ANALISIS DE DATOS EN ESPECTROSCOPIA GAMMA

El problema del análisis de un espectro de radiación gamma, establecido en la forma más general, consiste en la búsqueda de la transformación matemática inversa entre la distribución de altura de pulsos que conforma el espectro problema y el espectro de energía emitido por la fuente radiactiva. Una vez conocidas las distintas energías de emisión de una fuente radiactiva, la identificación de los radionúclidos que la forman es una tarea sencilla de búsqueda en una tabla de elementos y su energía o energías asociadas. En este capítulo se lleva a cabo una revisión de las técnicas aplicadas

en la actualidad al análisis del espectro de radiación obtenido en el analizador multicanal. El objetivo último de esta revisión es poner de manifiesto las ventajas e inconvenientes de todas y cada una de las distintas aproximaciones y retomar la que sea más conveniente a un doble propósito: primero, llegar a una técnica de análisis que excluya el concurso del operador humano tanto en la labor de mantenimiento del equipo de detección (autonomía) como en el de la interpretación de los datos del espectro medido (objetividad) y, segundo, concebir una arquitectura dedicada en base a procesadores de propósito específico que lleve a cabo la algoritmia apropiada. Esta última será desarrollada en posteriores capítulos.

2.1 LA FUNCION DE TRANSFERENCIA DEL DETECTOR

El espectro de emisión de una fuente radiactiva está formado por un conjunto de rayas de energía características de la emisión. El espectro observado en el analizador multicanal no es exactamente así. Toda observación experimental siempre sufre una degradación como resultado de que la resolución de la medida es finita y el ruido siempre hace acto de presencia. En espectroscopía gamma, la observación de las rayas espectrales se complica aún más. En el espectro medido se observa siempre una función más o menos continua que se denomina "fondo" y una serie de formas acampanadas que corresponden a las rayas espectrales de la emisión. Podemos decir que el sistema de detección actúa sobre

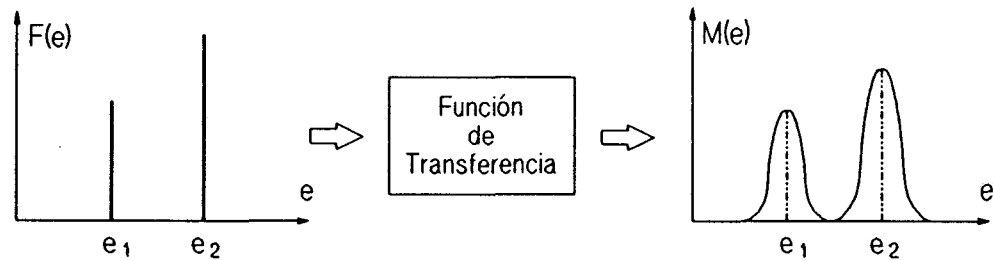


Fig. 2.1 El sistema de detección visto como una función de transferencia. Las rayas espectrales del espectro fuente se observan como formas acampanadas en el histograma de cuentas.

el espectro fuente como si se tratara de una función de transferencia que, aplicada sobre éste, produce el espectro observado tal como ilustra la Fig. 2.1. El objetivo de la espectroscopía gamma -y en general cualquier otra- es llevar a cabo el proceso inverso: llegar al espectro fuente conocido el espectro medido y conocida la función de transferencia. Antes de profundizar en este punto es necesario examinar los factores físicos que determinan ésta última.

2.1.1 EL ESPECTRO PROPORCIONADO POR EL DETECTOR

La función básica de un detector de radiación es producir un pulso eléctrico de amplitud proporcional a la energía E de cada partícula incidente en el mismo. Podemos representar gráficamente el número de pulsos obtenidos en la unidad de tiempo con amplitud comprendida entre E y $E + \Delta E$. El conjunto de datos obtenido se llama en estadística un histograma y en espectroscopía, lógicamente, un *espectro*. Supongamos que sobre el detector incide una radiación monoenergética de 2 MeV. El espectro proporcionado por un hipotético detector de resolución infinita es el que muestra la Fig. 2.2. Una porción determinada de fotones sufre interacción fotoeléctrica. El electrón desprendido será

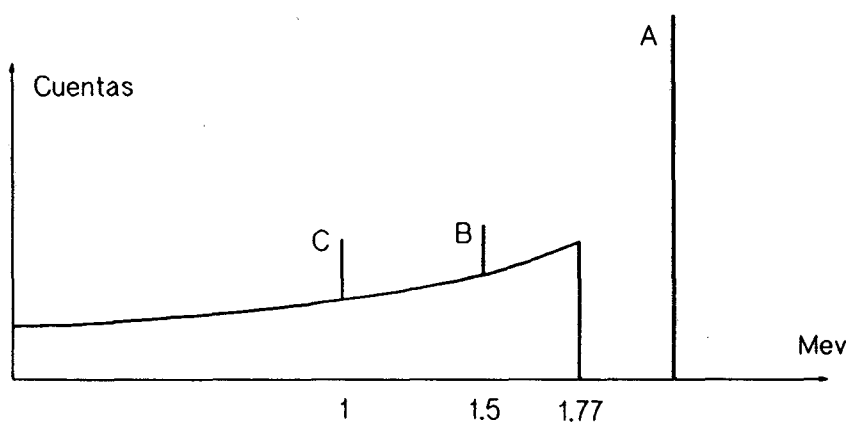


Fig. 2.2 Espectro ideal de una radiación monoenergética proporcionado por un detector. A: pico fotoeléctrico. B: pico de escape simple. C: pico de escape doble.

probablemente absorbido en el detector también por interacción fotoeléctrica. Como consecuencia, la amplitud del pulso resultante corresponderá a la energía de la radiación incidente. La acumulación en el tiempo del número de pulsos de esta amplitud se almacena en la raya espectral A, también llamada fotopico.

Otra proporción de fotones interacciona en modo Compton, dando origen a un electrón y a un fotón de menor energía. Este último tiene una cierta probabilidad de escapar del volumen del detector, mientras aquel perderá seguramente toda su energía en su interior. El electrón puede adquirir una energía cualquiera entre 0 y un valor máximo que depende de la energía del fotón incidente y que en este caso vale 1,77 MeV. La curva de energías provocada por este efecto viene indicada en la citada Fig. 2.2.

Finalmente, un fotón incidente puede dar lugar a la creación de un par electrón-positrón. La energía cinética de estas dos partículas será muy probablemente absorbida dentro del detector y en él tendrá lugar la aniquilación del positrón con la emisión simultánea en direcciones opuestas de dos fotones gamma de 0,501 MeV cada uno. Puede suceder que uno de estos dos fotones escape del cristal mientras que el otro sea absorbido, lo que dará lugar al pico B de la Fig. 2.2, llamado generalmente pico de escape simple y correspondiente a una energía de $2 \cdot 0,501$ o aproximadamente 1,5 MeV. Puede suceder que los dos fotones escapen sin interacción, lo que dará lugar al pico C o pico de escape doble, de una energía de aproximadamente 1 MeV.

También puede darse el caso de que ambos fotones de aniquilación hubiesen sido totalmente absorbidos en el cristal. Y, así mismo, en la interacción Compton el fotón gamma disperso puede a su vez ser totalmente absorbido, bien directamente por efecto fotoeléctrico o a través de nuevas interacciones Compton. En todos estos casos, la radiación gamma incidente pierde toda su energía dentro del detector en sucesos que se producen de forma prácticamente simultánea, resultando un pulso que contribuye al fotopico. La probabilidad relativa de cada uno de estos eventos depende de la energía de la partícula y del tamaño del detector. De la discusión anterior se deduce fácilmente que cuanto mayor sea en tamaño del detector mayor es la probabilidad de interacción fotoeléctrica y menor la contribución del fondo Compton al espectro final, lo que redundará en una caracterización

más clara de la radiación incidente.

2.1.2 EL ESPECTRO OBSERVADO EN EL ANALIZADOR MULTICANAL

El espectro proporcionado por el detector no resulta en ningún caso tan sencillo como muestra la Fig. 2.2. En lugar de líneas, en el analizador multicanal se obtiene una distribución parecida a una curva gaussiana de mayor o menor anchura (Fig. 1.5) en función de la calidad del equipo completo de detección, amplificación, clasificación y registro de pulsos. En última instancia, la medida de la amplitud de un impulso supone el recuento del número de cargas eléctricas que lo constituyen. La dispersión de este número viene dada por su desviación típica, que en primera aproximación puede considerarse igual a su raíz cuadrada. Así, cuanto más energética es la radiación incidente, mayor es la anchura del fotopico correspondiente. El ruido intrínseco de la electrónica que procesará el pulso contribuirá así mismo a aumentar la anchura del pico. Así, puede decirse que la calidad de un detector se caracteriza por la anchura de la gaussiana a la que da lugar una fuente monoenergética. Esta anchura, cuando se mide en la mitad de la altura de la curva acampanada se llama Γ o FWHM (full width at half maximum), y mide la resolución del detector para cada energía en términos de la razón:

$$R(E) = \frac{\Gamma}{E} \tag{2.1}$$

En cuanto al espectro completo, la resolución finita de un detector real conduce a espectros reales que semejan un suavizamiento del espectro teórico como ya comentamos. Véase la Fig. 2.3.

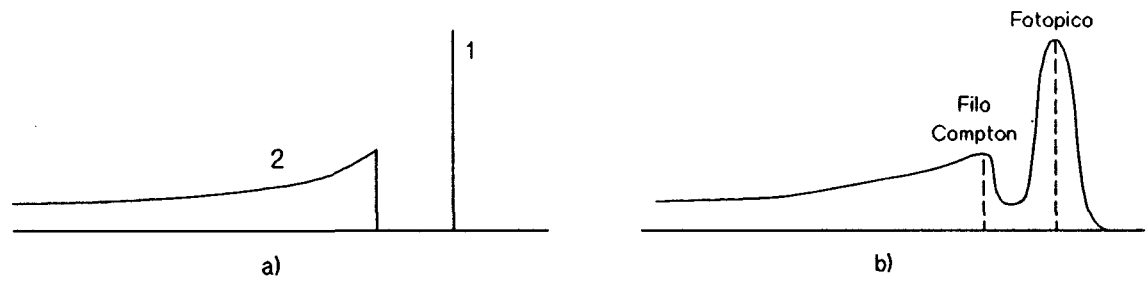


Fig. 2.3 Distribución de amplitud de pulsos de una fuente gamma monoenergética. (a) Resolución infinita. (b) Resolución finita.

2.2 LA DECONVOLUCION DEL ESPECTRO OBSERVADO

Ahora sabemos que dependiendo de la energía de la partícula incidente y del tipo de detector, una fuente monoenergética produce una distribución de amplitud de pulsos que tiene una forma más bien complicada. Podemos decir que, aunque el detector sea expuesto a una fuente radiactiva de una única energía, la distribución de amplitud del pulso se extiende sobre muchos canales del analizador (Fig. 2.3). El número relativo de pulsos de un determinado canal representa la probabilidad de que un rayo gamma de la fuente produzca un pulso en ese canal. Esta probabilidad es lo que se llama la *función de respuesta* o *función de resolución de energía* del sistema de detección. Así, se define $H(e, E)dE$ como la probabilidad de que una partícula emitida por una fuente de energía comprendida entre E y $E+dE$ sea registrada como de energía e . Dada esta definición, el espectro fuente y el espectro medido vienen relacionados por la ecuación:

$$m(e) = \int_0^{+\infty} H(e, E) f(E) dE . \tag{2.2}$$

Esta es una ecuación integral de primera clase, llamada ecuación de Fredholm ([91]), en

la que todas las funciones son no negativas y donde la incógnita es el espectro fuente $f(E)$. Su resolución se llama desdoblamiento o deconvolución. Es bien conocido en este contexto ([5], [14], [62], [69], [98], [99]) que el problema de la deconvolución del espectro medido es a menudo un problema mal planteado en el sentido de que factores como el ruido y pequeños cambios observados en la respuesta del detector $m(e)$ suponen grandes cambios en el espectro fuente $f(E)$. Antes de apuntar métodos de solución, es importante observar que el espectrómetro no mide $m(e)$, sino la cantidad

$$m_i = \int_{e_i}^{e_{i+1}} m(e) de , \quad 1 \leq i \leq n \quad (2.3)$$

donde $e_{i+1} - e_i = \Delta e_i$ es el intervalo de energía asociado a cada canal del espectrómetro. Dado que la ecuación (2.2) puede escribirse como

$$m(e) = \sum_{j=1}^n \int_{e_j}^{e_{j+1}} H(e, E) f(E) dE . \quad (2.4)$$

Sustituyendo (2.4) en (2.3) tenemos que

$$m_i = \sum_{j=1}^n \int_{e_i}^{e_{i+1}} de \int_{e_j}^{e_{j+1}} H(e, E) f(E) dE . \quad (2.5)$$

Este sistema de ecuaciones es aún exacto y uno de los métodos utilizados en la práctica para resolverlo es aproximar (2.5) a la ecuación

$$m_i = \sum_{j=1}^n H_{i,j} f_j , \quad 1 \leq i \leq n$$

donde

$$H_{ij} = \int_{e_i}^{e_{i+1}} H(e, E_j) de$$

y

$$f_j = \int_{e_j}^{e_{j+1}} f(E) dE .$$

En notación matricial:

$$m = Hf$$

donde el vector f debe ser interpretado como la aproximación discreta al espectro de emisión $f(E)$ dada por (2.4) y la matriz H como la aproximación discreta de la función de densidad de probabilidad $H(e, E)$ dada por (2.3). La referencia [62] sugiere aproximaciones de resolución continuas, pero el método generalmente adoptado es el basado en el conocimiento de la matriz inversa de la función respuesta:

$$f = h^{-1} m$$

Este método de desdoblamiento puramente algebraico tiene dos inconvenientes fundamentales: uno, la determinación de la matriz respuesta y, dos, el mal condicionamiento de la misma ([14], [82] y [98]) que da lugar a grandes errores en la determinación de la inversa. La matriz respuesta se halla excitando el detector con una fuente monoenergética que barre todo el rango de interés con un incremento energético dado ([7]). Conviene decir que no es un problema trivial el conseguir una fuente monoenergética pura ([39]). Una vez disponible este conjunto de respuestas empíricas, se interpola mediante técnicas apropiadas ([7], [34], [54] y [90]) para obtener el resto. Otra aproximación para obtener la matriz respuesta es el uso del métodos de Montecarlo ([39]). Estos están basados en simulaciones en computador de los procesos de interacción físicos antes descritos en el detector. La vía intermedia es la semiempírica, donde las respuestas Montecarlo vienen parametrizadas

([59]) y donde los parámetros se optimizan en función de las condiciones experimentales a base de ajustes sobre muestras de calibración. La Fig. 2.4 muestra una matriz respuesta típica de un detector de INa. Como puede apreciarse, las respuestas cercanas son muy parecidas, lo cual da lugar a filas prácticamente colineales en la matriz respuesta. Esta situación provoca un estado que ronda la singularidad de la misma por lo que surge la inestabilidad numérica en la construcción de la inversa. La técnica de interpolación utilizada en la referencia [90] se basa en este caso en un estudio de los componentes principales de la matriz respuesta ([90]). La consecuencia de todas las aproximaciones adoptadas en el proceso conducen a que el espectro fuente estimado no esté formado por rayas, sino por formas acampanadas. La Fig. 2.5 muestra el resultado de aplicar este tipo de deconvolución al espectro del isótopo ^{24}Na medido en un detector de INa ([39]) y un posterior suavizado. En esta ocasión se utilizó el programa comercial FORIST que utiliza la técnica antes descrita. Una revisión descriptiva de este tipo de programas de deconvolución se da en la referencia [83]. Como puede apreciarse, una vez hecha la deconvolución es necesario abordar un problema adicional que es el análisis de un espectro de campanas a través de cualquier técnica de búsqueda de picos. Esta es una tercera desventaja de este método de análisis, que se suma a las dos anteriores y permite concluir que el método de la matriz respuesta no es adecuado a un tratamiento automático del análisis de compuestos gamma.

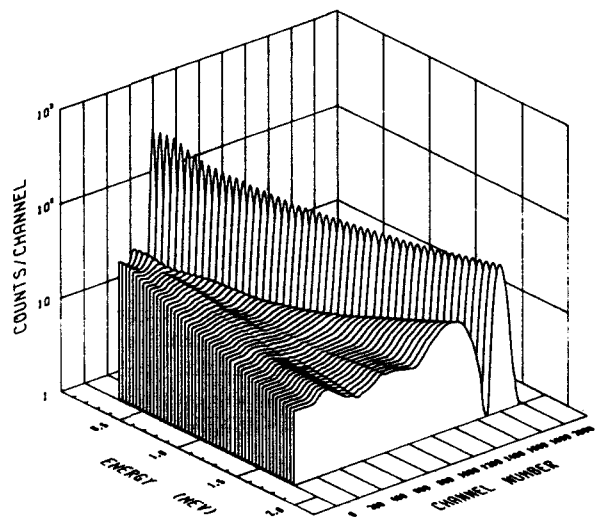


Fig. 11. Three-dimensional view of the response curves generated at equally spaced energies.

Fig. 2.4 Representación tridimensional de la matriz de curvas de respuesta a fuentes monoenergéticas. Tomada de [90].

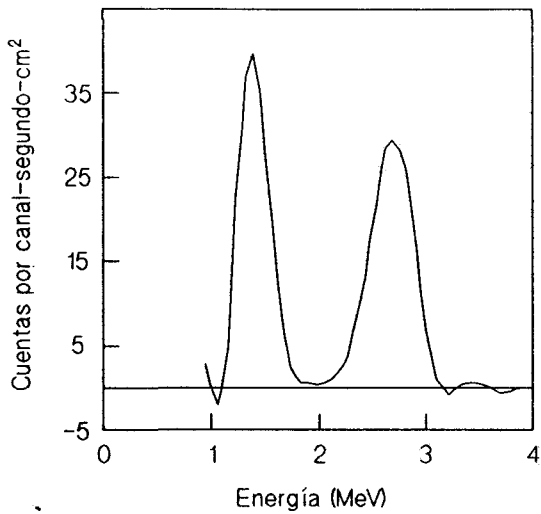


Fig. 2.5 Espectro de ^{24}Na -de dos rayas de emisión igualmente intensas a 1.37 y 2.75 Mev- desdoblado con el código comercial FORIST ([39]).

2.3 MEJORA DE RESOLUCION DEL ESPECTRO MEDIDO

La manera más directa de decidir cuáles son los radionúclidos que forman un espectro problema es detectar en éste las posiciones de todos y cada uno de los picos presentes en el mismo ([56]). La debilidad de esta técnica estriba en que muchas veces el equipo de medida no es capaz de resolver picos de energías similares. Como consecuencia, estos se muestran solapados unos con otros en el espectro, formando estructuras complejas llamadas dobletes, tripletes, etc, que enmascaran los picos formantes tal como muestra la Fig. 2.6b. Como puede observarse, la detección de estos picos no es tarea fácil y ello ha dado lugar a la búsqueda de técnicas que consigan mejorar la calidad del espectro obtenido a través de alguna transformación del mismo (Fig. 2.6c).

Desde principios de los años sesenta han ido apareciendo una serie de trabajos ([9], [40], [85] y [86]) acerca de la mejora de resolución de los picos presentes en un espectro de radiación. Según vimos en el anterior apartado, la operación de un espectróme-

tro viene caracterizada por la ecuación (2.2). La hipótesis fundamental de las técnicas utilizadas en las referencias anteriores consiste en modelar $H(e,E)$ como un sistema lineal invariante $h(e - E)$. Bajo esta suposición, la ecuación (2.2) se convierte en una ecuación de convolución ([70])

$$m = h * f, \quad (2.6)$$

por lo que a esta aproximación se le ha llamado deconvolución de Fourier. La ecuación anterior, expresada en el dominio de la frecuencia, se transforma en la ecuación $M = HF$. Podemos estimar H como la transformada de algún pico bien definido en un espectro dado. Así, por simple división, es posible calcular M y así mismo m mediante una transformación inversa. Esta aparente simplicidad presenta una debilidad fundamental que es su fuerte sensibilidad al ruido en el espectro medido m . Los valores de H se aproximan

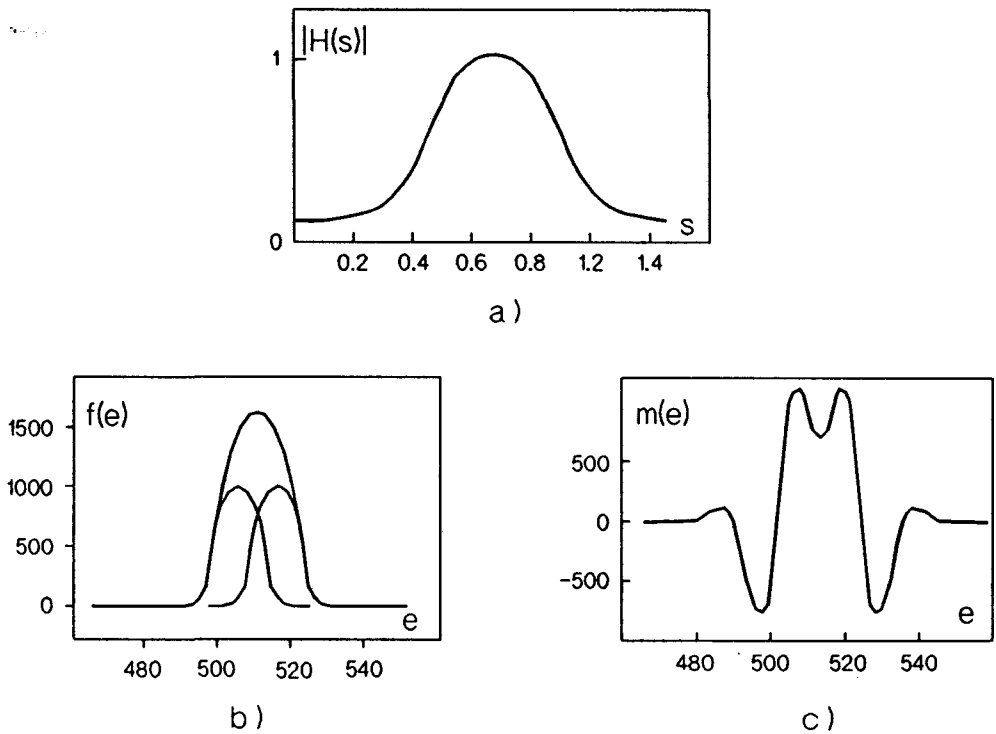


Fig. 2.6 (a) Filtro $H(e,E) = h(e-E)$ en el dominio de la frecuencia. (b) Superposición de dos picos gaussianos en el espectro medido. (c) Efecto de la función de resolución sobre el doblete. ([9]).

a cero a medida que aumenta la frecuencia debido a la suavidad de la función respuesta mientras que los de M no lo hacen debido al ruido. Por este motivo se produce una fuerte amplificación del ruido en la transformación inversa ([9]), pues la división M/H resulta altamente inexacta a altas frecuencias. La atenuación del ruido en los datos es, por tanto, un imperativo, por lo que un tratamiento previo de suavizado ([21], [22], [28], [45], [49], [79]) es condición "sine qua non" en la aplicación de la técnica, operación que no traslada la posición del centroide del pico si se lleva a cabo de forma apropiada ([9], [41]).

Por otra parte, los espectros de radiación presentan una dificultad añadida, y es que las técnicas de mejora de resolución requieren que todas las líneas adopten la misma forma y anchura, que es el modelo de respuesta al impulso característica del filtro modelado. Esta suposición no es en rigor válida en el espectro gamma, donde la anchura del pico crece con el número de canal. En todo caso, podría intentarse una aproximación por tramos de canales en los cuales la variación del modelo sea presumiblemente pequeña, como en los dobletes. Este es el caso del trabajo [63], referido a espectros de rayos X. Cuando no se da con el modelo adecuado, en el espectro fuente estimado aparecen picos espúreos ([41]). Según se desprende de las referencias citadas, la contribución más importante de esta técnica ha sido la de resolver dobletes en espectros procedentes de detectores de INa(Tl) ([9], [41]) tal y como se aprecia en Fig. 2.6.

En la ecuación (2.6) se asume tácitamente que el espectro medido m es un espectro de picos y, por lo tanto, está libre de cuentas de fondo Compton. Como esta suposición no se cumple en una medida real, antes de aplicar el modelo de convolución se debe sustraer previamente la contribución del fondo al espectro medido. Ya vimos que el fondo es el resultado de procesos físicos complejos y por esta razón no es posible dar de él una descripción analítica. Como consecuencia, hay que estimarlo directamente a partir del espectro medido. Una aproximación muy utilizada ([41], [46] y [63]) consiste en representar el fondo mediante una función suave que conecta todos los mínimos del espectro, entendido este como una función continua, una vez aplicado al mismo un proceso de suavizado. El método de interpolación en "splines" cúbicos ([61]) es el utilizado en la estimación de esta función en las referencias [46] y [63]. De todos modos, el resultado de estas técnicas de estimación de fondo es siempre subjetivo y cuestionable y, desde luego, tanto más exacto

cuanto mejor es la definición de los picos.

Técnicas más sofisticadas ([44], [47], [84] y [98]) de mejora de resolución de picos han aparecido en la literatura de los últimos años. Su falta de madurez, no obstante, y la ausencia de aplicaciones prácticas conocidas hace que se evite aquí cualquier comentario sobre las mismas.

2.4 BUSQUEDA DE PICOS

El campo de la espectroscopía gamma dio un gran paso hacia adelante en 1964 como consecuencia de la invención y desarrollo del detector de semiconductor de germanio. Su alta resolución simplificaba enormemente el análisis y permitía la detección de líneas muy débiles en la presencia de otras muy intensas y de energía muy cercana. Desde entonces, estos detectores y las técnicas de análisis asociadas a ellos, basadas en búsqueda de picos, se han multiplicado en la literatura. Aunque existen otros métodos ([27], [31], [51], [94]) para caracterizar un pico o abordar su estudio, el más utilizado consiste en ajustar la región del espectro donde se ha detectado un pico mediante una función analítica, tratando de minimizar el error cuadrático con respecto a un vector $a = [a_1, a_2, \dots, a_N]^T$ de parámetros libres:

$$e(a) = \sum_i W_i [y_i - f(i, a)]^2 \quad (2.7)$$

donde W_i es un factor de peso que suele tomarse como la inversa de y_i , siendo éste el número de cuentas acumuladas en el canal i . A este problema se le conoce en optimización como cuadrático no lineal sin restricciones ([6]).

2.4.1 DESCRIPCION FUNCIONAL DEL PICO

La función f que describe el espectro en la vecindad de un pico se modela arbitrariamente, ya que es impracticable derivarla teóricamente en rigor ([51], [75]) y ha sido dedicado mucho esfuerzo a buscar una función apropiada. Casi podríamos decir que cada nueva contribución en este sentido propone su propia función de ajuste, a veces con excesiva inclinación a aumentar el número de parámetros libres ([16]), con los consiguientes problemas numéricos que ello acarrea ([35]). Las referencias [36] y [57] constituyen un buen compendio de las más influyentes. Esta proliferación de métodos y las dificultades que entraña el compararlos ha llevado a distintos organismos internacionales [IAEA (Organización Internacional de la Energía Atómica, IEC (Comisión Internacional Electroatómica), etc] a la elaboración de un conjunto de espectros de Ge(Li) sobre los cuales cada método particular puede informar de los resultados obtenidos al analizarlos ([48], [65]). Si bien no todos los modelos la siguen ([76]), una de las funciones propuestas que ha gozado de mayor aceptación es la siguiente ([13], [20], [77]):

$$f(i, a) = G(i, a) + D(i, a) + S(i, a) + B(i, a) \quad (2.8)$$

donde

$$G(i, a) = a_1 + e^{\frac{-(i-a_2)^2}{2a_3^2}}, \quad (2.9)$$

$$D(i, a) = \begin{cases} a_3 e^{\frac{i-a_4}{a_5}}, & i \leq i_1 \\ a_3 e^{\frac{-(i-a_4)^2}{2a_6^2}}, & i \geq i_1, \end{cases} \quad (2.10)$$

donde a su vez $i_1 = a_4 - 2a_6^2/a_5$,

$$S(i, a) = \frac{a_7}{2} [1 - \operatorname{erf}(\frac{a_2 - i}{\sigma\sqrt{2}})] \quad (2.11)$$

y

$$B(i, a) = a_8 i + a_9 \quad (2.12)$$

todas ellas representadas en la Fig. 2.7. La elección de estas cuatro funciones se debe a que invariablemente un pico simple puede caracterizarse como una gaussiana distorsionada de alguna manera y superimpuesta sobre un fondo polinomial ([93]). La gaussiana $G(i)$ refleja la naturaleza estadística del ruido de fondo electrónico y el proceso de recolección de carga en el semiconductor como ya se discutió en el apartado 1.2. Los picos experimentales son, sin embargo, asimétricos. Esto requiere que el lado de menor energía de la gaussiana sea distorsionado con una cola exponencial $D(i)$ y una función escalón $S(i)$. Estos componentes de cola y escalón son atribuidos a una recolección de carga incompleta debida a fenómenos de atrapamiento, recombinación y al escape de electrones de la región sensible del detector ([13], [20], [43], [77]).

Si bien la resolución de los espectrómetros de Ge(Li) permite una definición extraordinariamente mayor que los detectores de INa(Tl), el problema de los multipletes no resueltos permanece. Este problema se aborda en la práctica replicando las ecuaciones (2.9),

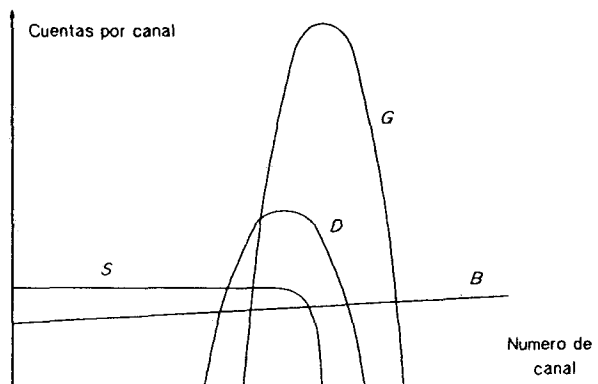


Fig. 2.7 Componentes analíticos aditivos de un pico simple.

(2.10) y (2.11), una para cada pico ([20], [37], [52]). De esta forma, el número de parámetros a determinar prácticamente se multiplica por el número de picos en el multiplete. Esta es la situación de análisis más conflictiva y de mayor tendencia al error, por lo que la mayoría de programas comerciales adoptan un modo interactivo con el usuario ([50], [55], [89]), siendo escasos los intentos de análisis totalmente automático ([56], [77], [87]). Es el usuario quien determina los límites de la región que contiene el multiplete y toma alguna decisión a cerca de eliminar o introducir algún pico para explicar el multiplete ([94]). En general, puede decirse que la cantidad de información de control proporcionada por el usuario aumenta con la complejidad del encaje ([35]). Por supuesto, esto introduce un factor de subjetividad en el análisis. La referencia [55] proporciona varios criterios para detectar multipletes en picos no resueltos por la rutina de búsqueda. La referencia [15] compara cuatro códigos comerciales a la hora de determinar la concentración más baja posible de un componente en una muestra problema.

2.4.2 LA DETECCION DEL PICO

La primera etapa de un análisis de picos consiste en recorrer el espectro desde un extremo energético al otro detectando intervalos de canales en los que es verosímil la presencia de una estructura de picos, bien un pico único, bien un doblete, triplete, etc. Generalmente son dos los tipos de errores encontrados en el proceso ([33]). Primero, fluctuaciones estadísticas del fondo se identifican erróneamente como fotopicos. Segundo, y al contrario, las fluctuaciones del fondo oscurecen los fotopicos. Se hace, pues, necesario suavizar el espectro para eliminar en lo posible las fluctuaciones más severas ([79]). El suavizado distorsiona el pico en relación inversa al número de canales sobre los que se extiende.

Una de las técnicas de búsqueda de picos más influyentes es la de las diferencias segundas ([30], [55], [56], [66]). Se fundamenta sobre un modelo de espectro entendido como la superposición de picos sobre una función suave y continua -el fondo-. De esta forma, la derivada segunda respecto al número de canal es prácticamente cero, por lo que se decide la presencia de un pico en un intervalo dado cuando la derivada segunda no se

anula o supera un determinado umbral. La Fig. 2.8 ilustra la idea. A este respecto, distintos criterios de discriminación son citados en [55]. El canal que ofrece el valor máximo en el intervalo se toma como centroide del pico, bien tentativo, como es usual para afinar después en la rutina de ajuste, o bien definitivo ([50]). La debilidad del método puede encontrarse en la referencia [87]. A pesar de ellas, la técnica de las diferencias segundas es utilizada en códigos comerciales, SAMPO ([50] y [77]) e HIPERMET ([74]) entre otros, aunque no es la única ([10], [12] y [17]), ya que la literatura es abundante en esta materia. El código comercial GAUSS ([37]), uno de los más populares, usa una simple técnica de correlación del espectro problema con una onda cuadrada. La referencia [53] proporciona un estudio comparativo de todas ellas.

Un intento de simplicidad en los procedimientos, sin requerimientos de ajuste de ningún tipo es el del trabajo [66]. En él, la determinación y caracterización del pico se realiza en base a la primera y segunda derivada del histograma. El precio pagado por esta sencillez en el método es su incapacidad para el tratamiento de multipletes.

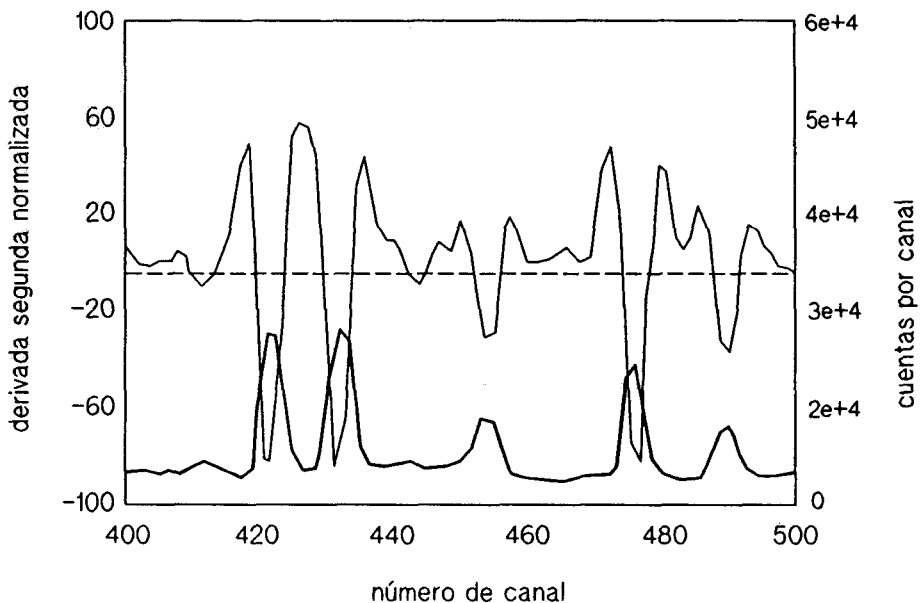


Fig. 2.8 Parte de un espectro gamma junto a su derivada segunda normalizada. Umbral de decisión a -5. Tomada de [55].

2.5 TECNICAS DE AJUSTE

El problema fundamental del análisis de picos es la determinación de un conjunto de parámetros de una función gaussiana o conjunto de ellas más o menos modificadas que mejor ajustan el histograma de cuentas en la región de interés (Fig. 2.9). El planteamiento matemático del problema citado viene dado por la minimización de la ecuación (2.7) con respecto al vector de parámetros a . Ya que la función $e(a)$ es una suma de cuadrados de funciones no lineales en a , las técnicas aplicadas son conocidas bajo la denominación de mínimos cuadrados no lineales ([6], [11], [24]-[26]). No existe en este contexto una solución dada de forma explícita, por lo que se hace necesaria la adopción de métodos iterativos. El problema de estos métodos estriba en que buscan el mínimo local más cercano a la configuración inicial (ver, por ejemplo [13]) y no el mínimo absoluto. En el contexto general de problemas de optimización no lineal ([19]) determinar este punto de partida cercano al óptimo es una condición imprescindible en la aplicación de estos métodos y el caso que nos ocupa no es una excepción. Tanto es así que, junto a la elección de las funciones de ajuste, la elección de estos valores iniciales es el problema más complejo encontrado en el análisis de picos de un espectro de radiación gamma ([89]). La forma de llevar a cabo la estimación inicial del vector de parámetros a varía de método en método ([92]) ya que depende de la función de ajuste escogida.

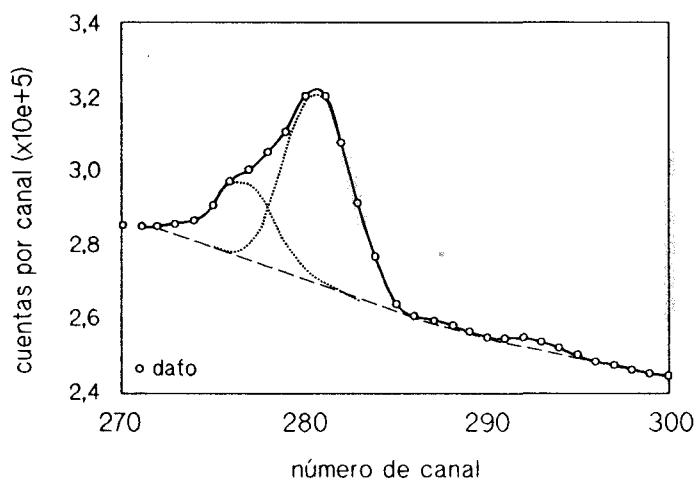


Fig. 2.9 Ajuste de un doblete mediante dos gaussianas y un polinomio para modelar el fondo bajo el pico. Tomada de [55].

Algunos métodos ([10], [50], [89]) utilizan el concepto de *pico de calibración* para reducir el número de parámetros a optimizar. La función de ajuste puede escogerse de tal forma que los parámetros varíen suavemente con la energía. Los estos pueden ser determinados separadamente para algunos picos aislados e interpolados para el resto. El código SAMPO utiliza dos parámetros para describir el fondo bajo el pico y cinco para el pico mismo. De estos últimos, tres de ellos son función de la energía. Esta reducción de parámetros tiene como consecuencia una convergencia más rápida y segura. La crítica de los picos de calibración se encuentra en [74]. Los autores argumentan que es necesaria una buena calibración y mantener el sistema muy estable o realizar una calibración para cada espectro medido, con la dificultad añadida de que a veces es difícil elegir buenos picos. Por otra parte, a partir de los 2 MeV la técnica de interpolación pierde gran parte de su validez, ya que la forma de los picos cambia de forma sustancial. Los códigos HYPERMET y GAUSS no utilizan esta técnica de parámetros interpolados.

2.5.1 TIEMPOS Y RECURSOS DE COMPUTACION

El "leit-motiv" de la relativamente reciente referencia [88] es la reducción de los tiempos y recursos de computación de las técnicas de ajuste de picos. La aproximación adoptada es no utilizar ajustes de mínimos cuadrados no lineales, sino lineales, que llevan a la resolución de un sistema lineal convencional. El uso del método de Gauss-Seidel propicia, según el autor, el uso de *aritmética entera*. Con el coste en exactitud que supone un modelo más simple, se reducen drásticamente los tiempos de computación. Según esta referencia, la descomposición de un multiplete de cuatro picos en un i8088 a 4 MHz lleva entre uno y dos segundos, mientras que la versión de SAMPO80 sobre un minicomputador tardaría 60 segundos, diez segundos sobre un VAX 11/780 y 2 segundos sobre un VAX 8650. La velocidad del código se acelera en un factor de 25 en un i80386 aprovechando al máximo la multiplicación y división entera de 32 bits de este micro. De esta breve discusión parece desprenderse que el transporte de los algoritmos de ajuste a sistemas de computación reducidos y por tanto susceptibles de ser portables a entornos ya no ajenos al laboratorio sino al del análisis de datos "in situ" y tiempo real pasan por la *simplicidad de la algoritmia y la aritmética*.

2.6 TECNICAS DE REGRESION MULTILINEAL

El análisis de un espectro de radiación mediante la identificación de fotopicos es aceptable para detectores de alta resolución. Sin embargo, los fotopicos y otras características espectrales no pueden ser observados con otro tipo de detectores como los de centelleo o semiconductores de CdTe y HgI₂ en energías mayores del MeV ([23], [72]). Por esta razón, el análisis de este tipo de datos requiere el desdoblamiento del espectro completo, bien mediante las técnicas examinadas en el apartado 2, bien mediante ajuste del espectro problema a partir de una base de referencias aplicando regresión lineal ([2], [8], [18], [32], [34], [38], [64], [67], [68], [71], [78], [80], [81], [95], [96], [97]). Estas últimas son examinadas a continuación.

2.6.1 FORMACION DEL ESPECTRO. CONCEPTOS FUNDAMENTALES

Consideremos una fuente radiactiva formada por un isótopo puro j situada en las cercanías de un detector dotado de un analizador multicanal de N canales. Durante un tiempo t_j recojemos S_{ij} cuentas en el canal i . Estas cuentas son atribuibles parte a la radiación misma y parte al ruido oscuro del detector, que en adelante denominaremos el fondo. Antes o después del periodo de exposición de la fuente j se recoge una muestra del fondo F_{ij} durante un periodo t_j^F . Se define la tasa de conteo del elemento j sobre el canal i como:

$$\chi_{ij} = (S_{ij}/t_j) - (F_{ij}/t_j^F)$$

Supongamos la emisión de una fuente compuesta por K isótopos elementales de modo que $1 \leq j \leq K$, cada uno de ellos con una tasa de conteo χ_{ij} . Se cumple entonces que, para cada canal, la tasa de conteo de la muestra t_i es igual a la suma de las tasas de

conteo de los isótopos que la componen:

$$l_i = \sum_{j=1}^K x_{ij}$$

La consecuencia inmediata de esta ley física es el *principio de superposición*, que establece que el número de cuentas y_i de una fuente compuesta y desconocida y recogidas en el canal i del analizador puede ser expresado como una combinación lineal del número de cuentas acumuladas por cada componente o referencia x_{ij} , $0 \leq j \leq K$, siendo cada coeficiente la contribución relativa de cada referencia a la mezcla. Así, para un analizador de N canales y K referencias más un vector de fondo agrupados en una matriz X :

$$y_i = \sum_{j=0}^K c_j X_{ij}, \quad 0 \leq i \leq N-1. \quad (2.13)$$

La ecuación (2.13) constituye un sistema lineal donde la incógnita es el vector de contribuciones. Como es sabido, admite solución para $N \leq K+1$. Generalmente el número de canales es bastante mayor que el elementos puros (en adelante referencias). Así podemos disponer de 2048 canales y tener una base 40 referencias. Este tipo de sistemas se llaman sobredimensionados porque el número de observaciones (ecuaciones) supera al de parámetros que han de ser estimados (incógnitas). Dado que esta característica impide una solución exacta, los esfuerzos se dirigen hacia la búsqueda de la mejor aproximación. La redundancia de ecuaciones es muy útil. Su función es suavizar las fluctuaciones estadísticas en la formación del espectro problema y que de alguna manera distorsionan la estimación de las contribuciones. Ya que la formación de un espectro de cuentas es en sí mismo un proceso estadístico, una formulación más rigurosa del problema exige la introducción de conceptos y términos de esta naturaleza.

2.6.2 EL MODELO DE REGRESION LINEAL MULTIPLE

El modelo estadístico de regresión lineal múltiple utiliza observaciones experimentales para estimar un vector de parámetros ([18], [73]). Sea la matriz X de K variables vectoriales independientes deterministas (x_1, x_2, \dots, x_K) de dimensión N y sea la variable aleatoria vectorial dependiente y de dimensión N . Se establecen las siguientes hipótesis sobre su distribución:

- 1.- Para cada X , la distribución de y es normal con media
$$E[y/X] = c_1x_1 + c_2x_2 + \dots + c_Kx_K.$$
- 2.- La varianza es constante e independiente de X : $\text{Var}[y_i/X] = \sigma^2, \quad 1 \leq i \leq N$
- 3.- Las variables y_i son independientes entre sí.
- 4.- El número de variables independientes K es menor que el de observaciones N .
- 5.- Entre las x_i no existen relaciones lineales.

En estos términos, el conjunto de observaciones puede ser expresado como $y_i = c_1X_{i1} + \dots + c_KX_{iK} + u_i, \quad 1 \leq i \leq N$ o, en notación matricial $y = Xc + u$, siendo u una perturbación aleatoria sujeta a las siguientes hipótesis:

- 1.- u es normal multivariada con $E[u] = 0$.
- 2.- La matriz de covarianza de u , $E[uu^T] = \sigma^2I_N$, donde I_N representa la matriz identidad de orden N .

2.6.3 EL MODELO ESTADISTICO DEL ESPECTRO PROBLEMA

Aun replicando exactamente las condiciones experimentales en el conteo una fuente, el espectro obtenido siempre varía como consecuencia de la naturaleza estadística de la emisión de radiación y los procesos de recolección de carga en el detector. Al introducir en el modelo este elemento de incertidumbre, una descripción más adecuada de la ecuación (2.13) es la siguiente:

$$E[y] = Xc$$

En esta ecuación, y es un vector columna de variables aleatorias $(y_1, y_2, \dots, y_N)^T$ y $E[y] = (E[y_1], E[y_2], \dots, E[y_N])^T$ donde $E[y_i]$, $1 \leq i \leq N$ es el valor esperado de la variable aleatoria y_i . Cada y_i tiene una distribución de Poisson con media y varianza $E[y_i]$. Para una realización de y_i suficientemente grande (más de 50 cuentas según [71] y más de 20 según [91]), y una contribución del fondo pequeña, su distribución puede considerarse normal de media $E[y_i]$ e idéntica varianza $E[y_i]$ a partir del teorema central del límite ([18], [91]). Entonces, si y tiene una distribución normal multivariada, su función de densidad es:

$$f(y) = \left[\frac{1}{2\pi} \right]^{N/2} \Sigma^{-1} \exp\left(-\frac{1}{2}(y - Xc)^T \Sigma^{-1} (y - Xc)\right) \quad (2.14)$$

en la cual el vector de contribuciones c actúa como parámetro.

Una de las técnicas de estimación paramétrica utilizada en estadística es el método de máxima verosimilitud. Dada una realización y y una distribución conocida a priori salvo un conjunto o vector de parámetros, estima éste de tal forma que la función de densidad resultante evaluada en la muestra sea la mayor de todas ([58]). Maximizar (2.14) se traduce, por lo tanto, en el siguiente problema de optimización:

$$\underset{(c)}{\text{Minimizar}} \quad (y - Xc)^T \Sigma^{-1} (y - Xc) \quad (2.15)$$

La matriz Σ es la matriz de covarianzas de las variables aleatorias $(y_1, y_2, \dots, y_N)^T$.

Se define como:

$$\Sigma \equiv E[(y - E[y])(y - E[y])^T],$$

que en este caso es

$$\Sigma = E[(y - Xc)(y - Xc)^T].$$

Como vemos, el valor de la inversa de la matriz de covarianza depende del vector de contribuciones de una forma fuertemente no lineal, lo cual complica extraordinariamente la minimización de $(y - Xc)^T \Sigma^{-1} (y - Xc)$ respecto a c , por lo que es necesario estimarla a priori a partir del espectro problema y . La práctica más común es considerar que las variables aleatorias y_1, y_2, \dots, y_N son estadísticamente independientes, por lo que $\text{Cov}[y_i, y_j] = 0$ si $i \neq j$. Así, la matriz de covarianza Σ es diagonal, lo que supone que también lo es su inversa Σ^{-1} ([29]). De esta forma, el problema de minimización (2.16) se simplifica en el siguiente:

$$\underset{(c)}{\text{Minimizar}} \quad \sum_{i=1}^N \left[\frac{1}{y_i} \right] \left[y_i - \sum_{j=0}^K c_j X_{ij} \right]^2 \quad (2.16)$$

llamado método de los *mínimos cuadrados ponderados*, que conduce a dar mayor peso a los canales que se estima tienen menor varianza, precisamente aquellos con menor número de cuentas acumuladas. En la referencia [64] se argumenta que esta aproximación a la matriz de covarianza es adecuada, pero en la bibliografía se da algún trabajo que no la sigue ([78], [95]).

La solución al problema (2.16) se halla derivando respecto al vector de parámetros e igualando a cero, lo que nos lleva a la solución:

$$\mathbf{c} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{y}.$$

\mathbf{c} es un vector aleatorio que es función lineal del vector aleatorio \mathbf{y} . Es importante observar que su valor no cambia al multiplicar Σ por una constante, lo que se traduce en la práctica en la robustez del estimador frente a ruido gaussiano (error estadístico) en la formación del espectro. Por otra parte, si la distribución del vector \mathbf{y} es normal $\mathbf{y} \sim N(\mathbf{X}\mathbf{c}, \Sigma)$, entonces la distribución de \mathbf{c} también es normal ([18], [73]) $\mathbf{c} \sim N(\mathbf{c}, (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1})$. El elemento i de la diagonal de la matriz de covarianza $\mathbf{X}^T \Sigma^{-1} \mathbf{X}$ es la varianza del estimador c_i y su raíz cuadrada la desviación típica de la estimación, con lo cual la construcción de intervalos de confianza de la misma es inmediata.

2.6.3.1 BONDAD DEL ENCAJE

El modelo de formación del espectro de la ecuación (2.14) es correcto bajo dos supuestos fundamentales. Primero, la ausencia de derivas en el equipo de medida y, segundo, la matriz \mathbf{X} contiene las referencias de los isótopos que están presentes en el espectro problema \mathbf{y} . En este caso, la expresión que trata de minimizarse en la ecuación (2.16), $(\mathbf{y} - \mathbf{X}\mathbf{c})^T \Sigma^{-1} (\mathbf{y} - \mathbf{X}\mathbf{c})$, es una variable aleatoria que tiene una distribución χ^2 con $N-K$ grados de libertad y su valor esperado es $N-K$ ([38]). Si definimos $\chi_F^2 = \chi^2 / (N-K)$, el valor esperado de χ_F^2 es la unidad. A efectos prácticos, cuanto más cercano a la unidad esté este valor, más fiable es el análisis del espectro problema \mathbf{y} . Su valor aumenta si existen derivas instrumentales o el conjunto de referencias es insuficiente.

Desgraciadamente, el valor de la χ_F^2 no proporciona ninguna indicación acerca de la naturaleza del error del modelo ([1], [4]). La literatura ha tratado este problema ampliamente buscando indicadores más representativos de las fuentes de discrepancia, tanto

en el ámbito de ajuste de picos ([1], [3], [4], [42], [60]) como en el de formas espectrales completas ([71]). Sin embargo, de una forma u otra, casi todas son derivaciones del clásico test de la χ^2 más elementos añadidos "ad hoc" derivados de la experiencia.

2.7 BIBLIOGRAFIA

- [1] Aarnio, A. *et al.*, "A New Method for the Determination of Goodness-of-Fit in Gamma Spectroscopy", *Nuclear Instruments and Methods*, 184, (1981), pp. 478-492.
- [2] Antoon, M. K., *et al.*, "Least-squares Curve-fitting of Fourier Transform Infrared Spectra with Applications to Polymer Systems", *Applied Spectroscopy*, Vol. 31, No. 6, (1977), pp. 518-524.
- [3] Awaya, T., "A New Method For Curve Fitting To the Data With Low Statistics Not Using the χ^2 Method", *Nuclear Instruments and Methods*, 165, (1979), pp. 317-323.
- [4] Balian, H. G. *et* Eddy, N. W., "Figure-Of-Merit (FOM), an Improved Criterion Over the Normalized Chi-Squared Test For Assessing Goodness-Of-Fit Of Gamma-Ray Spectral Peaks", *Nuclear Instruments and Methods*, 145, (1977), pp. 389-395.
- [5] Belogorlov, E. A. *et* Zhigunov, V. P., "Interpretation of the Solution to the Inverse Problem for the Positive Function and Reconstruction of Neutron Spectra", *Nuclear Instruments and Methods A235*, (1985), pp. 146-163.
- [6] Bevington, P. R., "Data Reduction and Error Analysis in the Physical Sciences", McGraw-Hill, (1969).
- [7] Bishop, G. B. *et* Marafie, A. M., "Formation of a Response Matrix to Provide Benchmark Data for 6 MeV Gamma Photon Penetrations Using a NaI Spectrometer", *Nuclear Instruments and Methods*, 150, (1978), pp. 505-508.
- [8] Blackburn, J., "Computer Program for Multicomponent Spectrum Analysis Using Least-

Squares Method", *Analytical Chemistry*, 37, No. 8, July (1965), pp. 1000-1004.

- [9] Blinowska, K. J. *et* Wessner, E. F., "A Method of On-Line Spectra Evaluation by Means of a Small Computer Employing Fourier Transforms", *Nuclear Instruments and Methods*, 118, (1974), pp. 579-604.
- [10] Block, H. P., DeLange, J. C. *et* Schotman, J. W., "A New Peak Search Method for Automatic Spectrum Analysis Program", *Nuclear Instruments and Methods*, 128, (1975), pp. 545-556.
- [11] Brown, K. M., *et* Dennis, J. E., "Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximation", *Numer. Math*, 18, (1972), p. 289-297.
- [12] Bryant, W. F., *et al*, "Data-Blocking Cross-Correlation Peak Detection in Computerized Gas Chromatography-Mass Spectrometry", *Analytical Chemistry*, (1980), 52, pp. 38-43.
- [13] Campbell, J. L. *et* Jorch, H. H., "On the Analytic Fitting of Full Energy Peaks from Ge(Li) and Si(Li) Photon Detectors. II", *Nuclear Instruments and Methods*, 159, (1979), pp. 163-170.
- [14] Chambless, D. A. *et* Broadway, J. A., "Constrained Regularization in Nuclear Spectroscopy", *Nuclear Instruments and Methods*, 179, (1981), pp. 563-571.
- [15] Cline, J. E., "A Comparison of Detection-Limit Computations for Four Commercial Gamma-Ray Analysis Programs", *Nuclear Instruments and Methods*, A286, (1990), pp. 421-428.
- [16] Coldwell, R. L., "Iterative Codes for Fitting Complete Spectra", *Nuclear Instruments and Methods*, A242, (1986), pp. 445-461.
- [17] Connely, A. L. *et* Black, W. W., "Automatic Location and Area Determination of Photo-peaks", *Nuclear Instruments and Methods*, 82, (1970), pp. 141-148.
- [18] Covell, D. F. *et al.*, "Multiple Linear Regression Analysis of Scintillation Gamma-Ray Spectra: Theoretical and Practical Considerations", *Nuclear Instruments and Methods*, 80, (1970), pp. 55-60.

- [19] Demidovich, B. P., *et* Maron, I. A., "Cálculo numérico fundamental", Editorial Paraninfo, (1985).
- [20] Dojo, M., "A Shape Function of Photopeaks for Gamma-Ray Spectrum Analysis with Ge(Li) Detectors", *Nucler Instruments and Methods*, 115, (1974), pp. 425-429.
- [21] Edwards, T. H. *et* Willson, P. D., "Digital Smoothing of Spectra", *Applied Spectroscopy*, 28, (1974), pp. 541-545.
- [22] Enke, C. G. *et* Nieman, T. A., "Signal-To-Noise Ratio Enhancement By Least Squares Polynomial Smoothing", *Analytical Chemistry*, Vol. 48, No. 8, (1976), pp. 705-712.
- [23] Entine, G. *et al.*, "Survey of CdTe Nuclear Detector Applications", *Nuclear Instruments and Methods*, A283, (1989), pp. 282-290.
- [24] Fletcher, R., "A New Approach to Variable Metric Algorithms", *Computer J.*, 13, (1970), p. 317-322.
- [25] Fletcher, R., "A Modified Marquard Subroutine For Non-Linear Least Squares", Report AERE R-6799, Theoretical Physics Division, Atomic Energy Research Establishment Harwell, Berkshire, (1971).
- [26] Fletcher, R. *et* Powell, M. J. D., "A Rapidly Convergent Descent Method For Minimization", *The Computer Journal*, Vol. 6, (1963), pp. 163-168.
- [27] Gemperline, P. J. *et* Hamilton, J. C., "Conditions for Detecting Overlapped Peaks with Principal Components Analysis in Hyphenated Chromatographic Methods", *Analytical Chemistry*, 61, (1989), pp. 2240-2243.
- [28] Gorry, P. A., "General Least-Squares Smoothing and Differentiation By the Convolution (Savitzky-Golay) Method", 62, (1990), pp. 570-573.
- [29] Griffel, D. H., "Linear Algebra and its Applications", John Wiley & Sons, 1989.
- [30] Grushka, E. *et* Israeli, D., "Characterization of Overlapped Chromatographic Peaks by Their Second Derivative. The Limit of the Method", *Analytical Chemistry*, (1990), 62, pp. 717-721.

- [31] Grushka, E. *et* Monacelli, G. C., "Slope Analysis for Recognition and Characterization of Strongly Overlapped Chromatographic Peaks", *Analytical Chemistry*, Vol. 44, No. 3, March, (1972), pp. 484-489.
- [32] Haaland, D. M., "Quantitative Infrared Analysis of Borphosphosilicate Films Using Multivariate Statistical Methods", *Analytical Chemistry*, 60, (1988), pp. 1208-1217.
- [33] Head, J. H., "Minimum Detectable Area in Ge(Li) Spectra", *Nuclear Instruments and Methods*, 98, pp. 419-428, (1972).
- [34] Heath, R. L., "Data Anaysis Techniques for Gamma-Ray Scintillation Spectrometry", *Atomic Energie Comission*, Report IDO-16784, (1962).
- [35] Helmer, R. G. *et al.*, "Photopeaks Analisis Program for Photon Energy and Intensity Determinations", *Nuclear Instruments and Methods*, 57, (1967) pp. 46-57.
- [36] Helmer, R. G. *et* Lee, M. A., "Analytical Functions for Fitting Peaks from Ge Semiconductor Detectors", *Nuclear Instruments and Methods*, 178, (1980), pp. 499-512.
- [37] Helmer, R. G. *et* McCullagh, C. M., "GAUSS VII, a Computer Program for the Analysis of Gamma-Ray Spectra from Ge Semiconductor Spectrometers", *Nuclear Instruments and Methods*, 206, (1983), pp. 477-488.
- [38] Hogan, M. A. *et al.*, "Multiple Linear Regression Analysis of Scintillation Gamma-Ray Spectra: Automatic Candidate Selection", *Nuclear Instruments and Methods*, 80, (1970), pp. 61-68.
- [39] Ingersoll, D. T. *et* Wehring, B. W. "Gamma-Ray Pulse-Height Response of an NE-213 Scintillation Detector", *Nuclear Instruments and Methods*, 1977, pp. 551-561.
- [40] Inouye, T., "The Super Resolution of Gamma-Ray Spectrum", *Nuclear Instruments and Methods*, 30, 1964, pp. 224-228.
- [41] Inouye, T., "Aplication of Fourier Transforms to the Analysis of Spectral Data", *Nuclear Instruments and Methods*, 67, 1969, pp. 125-132.

- [42] Isozumi, Y., "Uncertainties In the χ^2 Fit For Nuclear Radiation Spectra", *Nuclear Instruments and Methods*, A235, (1985), pp. 164-173.
- [43] Jorch, H. H. *et* Campbell, J. L., "On the Analytical Fitting of Full Energy Peaks From Ge(Li) and Si(Li) Photon Detectors", *Nuclear Instruments and Methods*, 143, (1977), pp. 551-559.
- [44] Kauppinn, J. K., Moffatt, D. J., Hollberg, M. R. *et* Mantsch, H. H., "A New Line-Narrowing Procedure Based on Fourier Self-Deconvolution, Maximum Entropy, and Linear Prediction", *Applied Spectroscopy*, Vol. 45, No. 3, (1991), pp. 411-416.
- [45] Kawata, S. *et* Minami, S., "Adaptive Smoothing of Spectroscopic Data By a Linear Mean-Square Estimation", *Applied Spectroscopy*, Vol. 38, No. 1, (1984), pp. 49-58.
- [46] Kennett, T. J., Prestwich, W. V. *et* Tervo, R. J., "Automated Analysis for High Energy Gamma Ray Spectra", *Nuclear Instruments and Methods*, 1981, pp. 313-323.
- [47] Kennet, T. J. *et* Prestwich, W. V., "Incremental Deconvolution I. Algorithm development and assessment", *Nuclear Instrument and Methods*, 203, (1982), pp. 317-327.
- [48] Keyser, R. M., "Using Standard Spectra to Development and Test Gamma-Ray Analysis Software", *Nuclear Instruments and Methods*, A286, (1990), pp. 409-414.
- [49] Khan, A., "Problems of Smoothing and Differenciation of Data By Least Squares and Possible Solutions", *Analytical Chemistry*, 59, (1987), pp. 654-657.
- [50] Koskelo, M. J., Aarnio, P. A. *et* Routti, J. T., "SAMPO80: An Accurate Gamma Spectrum Analysis Method for Minicomputers", *Nuclear Instruments and Methods*, 190, (1981), pp. 89-99.
- [51] Kotka, L., "Determination of Peak Area", *Nuclear Instruments and Methods*, 112, (1973), pp. 245-251.
- [52] Lauterjung, J., Will, G. *et* Hinze, E., "A Fully Automatic Peak-Search Program for the Evaluation of Gauss-Shaped Diffraction Patterns", *Nuclear Instruments and Methods*, A239, (1985), pp. 281-287.

- [53] De Lotto, I. *et* Ghirardi, A., "Automatic Peak-Locating Techniques for Gamma-Ray Spectra", *Nuclear Instruments and Methods*, 143, (1977), pp. 617-620.
- [54] Luri, N. A., Harris, L. *et* Young, J. C., "Calculation of Gamma Ray Response Matrix for 5 cm NE-213 Organic Liquid Scintillator Detector", *Nuclear Instruments and Methods*, 129, (1975), pp. 543-555.
- [55] Machner, H., "Automatic Analysis of Gamma-Ray Spectra Including Multiplets", *Nuclear Instruments and Methods*, A258, (1987), pp. 246-249.
- [56] Mariscotti, M. A., "A Method for Automatic Identification of Peaks in the Presence of Background and its Application to Spectrum Analysis", *Nuclear Instruments and Methods*, 57, (1967), pp. 309-320.
- [57] McNelles, L. A. *et* Campbell, J. L., "Analytic Aproximations to Peak Shapes Produced by Ge(Li) and Si(Li) Spectrometers", *Nuclear Instruments and Methods*, 127, (1975), pp. 73-81.
- [58] Meyer, P. L., "Probabilidad y aplicaciones estadísticas", Fondo Educativo Interamericano, (1973).
- [59] Mitchel, D. J. *et al.*, "Gamma-Ray Respose Functions for Scintillation and Semiconductor Detectors", *Nuclear Instruments and Methods*, A276, (1989), pp. 547-556.
- [60] Misra, S. K. *et* Eddy, N. W., "IFOM, a Formula For Universal Assessment Of Goodness-Of-Fit Of Gamma Ray Spectra", *Nuclear Instruments and Methods*, 166, (1979), pp. 537-540.
- [61] Morton, W. T., "Knot Positions in Least-Squares Fitting of Data Using Cubic Splines", *Nuclear Instruments and Methods*, A272, (1988), pp. 861-865.
- [62] Multhei, H. N. *et* Schorr, B., "On an Iterative Method for Unfolding of Spectra", *Nuclear Instruments and Methods*, A257, (1987), pp. 371-377.
- [63] Nass, M. J., Lurio, A. *et* Ziegler, J. F., "Computer Analysis of X-Ray Spectra", *Nuclear Instruments and Methods*, 154, (1978), pp. 567-571.
- [64] Nicholson, W. L. *et al.*, "The Quantitative Analysis of Sets of Multicomponent Time

- Dependent Spectra from Decay of Radionuclides", *Nuclear Instruments and Methods*, 25, (1963), pp. 45-66.
- [65] Nielsen, S. P., "Quantitative Evaluation of Gamma-Spectrum Analysis Methods Using IAEA Test Spectra", *Nuclear Instruments and Methods*, 192, (1982), pp. 433-438.
- [66] Nyman, B., "Techniques for Fast Evaluations of Ge(Li)-Detector Gamma-Ray Spectra with Small Computers", *Nuclear Instruments and Methods*, 108, (1973), pp. 273-241.
- [67] Olmos, P., Díaz, J. C. *et al.*, "A New Approach To Automatic Radiation Spectrum Analysis", *IEEE Trans. on Nuclear Science*, Vol. 38, No. 4, August, (1991), pp. 971-975.
- [68] Olmos, P., Díaz, J. C. *et al.*, "Application of Neural Network Techniques In Gamma Spectroscopy", *Nuclear Instruments and Methods*, A312, (1992), pp. 167-173.
- [69] Onoda, T. *et* Sekimoto, H., "A neutron Spectrometry Unfolding Code Based on Quadratic Programming", *Nuclear Instruments and Methods*, A272, (1988), pp. 844-846.
- [70] Oppenheim, A. V., Willsky, A. S. *et* Young, I. T., "Signals and Systems", Prentice-Hall, (1983).
- [71] Parr, R. M. *et* Lucas, H. F., "A Rigorous Least Squares Analysis of Complex Gamma Ray Spectra with Partial Compensation for Instrumental Instability", *IEEE Trans. on Nuclear Science*, 349, June, (1964), pp. 349-357.
- [72] Patt, B. E. *et al.*, "Developments in Mercuric Iodide Gamma Ray Imaging", *Nuclear Instruments and Methods*, A283, (1989), pp. 215-219.
- [73] Peña, D., "Estadística: Modelos y Métodos. 2: Modelos Lineales y Series Temporales". Alianza Editorial, (1989).
- [74] Phillips, G. W. *et* Marlow, K. W., "Automatic Analysis of Gamma-Ray Spectra from Germanium Detectors", *Nuclear Instruments and Methods*, 137, (1976), pp. 525-536.
- [75] Quittner, P., "Peak Area Determination for Ge(Li) Detector Data", *Nuclear Instruments and Methods*, 76, (1969) pp. 115-124.

- [76] Robinson, D. C., "A Computer Programme for the Determination of Accurate Gamma Ray Line Intensities from Germanium Spectra", *Nuclear Instruments and Methods*, 78, (1970), pp. 120-124.
- [77] Routti, J. T. *et* Prussin, S. G., "Photopeak Method for the Computer Analysis of Gamma-Ray Spectra from Semiconductor Detectors", *Nuclear Instruments and Methods*, 72, (1969), pp. 125-142.
- [78] Salmon, L., "Analysis of Gamma-Ray Scintillation Spectra by the Method of Least Squares", *Nuclear Instruments and Methods*, 14, (1961), pp. 193-199.
- [79] Savitzky, A. *et* Golay, M. J., "Smoothing and Diferenciation of Data by Simplified Least Squares Procedures", *Analytical Chemistry*, 36, (1964), pp. 1627-1639.
- [80] Schonfeld, E. *et al.*, "Determination of Nuclide Concentrations in Solutions Containing Low Levels of Radiativity by Least-Squares Resolution of the Gamma-Ray Spectra", *Nuclear Instrumens and Methods*, 45, (1966), pp. 1-22.
- [81] Schonfeld, E., "ALPHA - a Computer Program for the Determination of Radioisotopes by Least-Squares Resolution of the Gamma-Ray Spectra", *Nuclear Instruments and Methods*, 42, (1966), pp. 213-218.
- [82] Sekimoto, H., "An Unfolding Method Leading to a Positive Solution Only", *Nuclear Instruments and Methods*, 288, (1984), pp. 129-132.
- [83] Sekimoto, H. *et* Yamamuro, N., "Unfolding Methods whith the Prior Knowledge and Integral Data of Neutron Spectrum", *Nuclear Science and Engineering*, 80, (1982), pp. 101-112.
- [84] Sivia, D. S., Vorderwisch, P. *et* Silver, R. N., "Deconvolution of Data from the Filter Difference Spectrometer: from Hardware to Maximum Entropy", *Nuclear Instrument and Methods*, A290, (1990), pp. 492-498.
- [85] Sjontoft, E. "A Straightforward Deconvolution Method for Use in Small Computers", *Nuclear Instruments and Methods*, 163, (1979), pp. 519-522.
- [86] Sjontoft, E. "Deconvolution when Only the Lower Order Moments of the Convolution

- Function are Known", *Nuclear Instruments and Methods*, 206, (1983), pp. 199-202.
- [87] Slavic, I. A., "Automatic Analysis of Gamma-Ray Spectra", *Nuclear Instruments and Methods*, 112, (1973), pp. 253-260.
- [88] Stockman, H. W., "Microprocessor-Based Reduction of Gamma Spectra: a Fast Method for Deconvolution of Overlapped Peaks", *Nuclear Instruments and Methods*, A274, (1989), pp. 314-318.
- [89] Tasaka, K., "Standard Spectrum Method for the Analysis of Gamma-Ray Spectra from Semiconductor Detectors", *Nuclear Instruments and Methods*, 109, (1973), pp. 547-556.
- [90] Tominaga, S., "An Estimation Method for Generating Scintillator Response Curves for Gamma-Ray Analysis", *Nuclear Instruments and Methods*, 205, (1983), pp. 485-493.
- [91] Tsoufanidis, N., "Measurement and Detection of Radiation", McGraw-Hill, 1983.
- [92] West, D. H. D., "Incremental Least-Squares and the Aproximate Separation of Exponentials", *Nuclear Instruments and Methods*, 205, (1976), pp. 137-143.
- [93] Welch, R. B. *et al.* "NEWFIT, a Computer Program for the Analysis of Alfa, X-Ray and Gamma-Ray Spectra", *Nuclear Instruments and Methods*, A269, (1988), pp. 615-622.
- [94] Wythoff, B. J. *et al.*, "Spectral Peak Verification and Recognition Using a Multilayered Neural Network", *Analytical Chemistry*, 62, (1990), pp. 2702-2709.
- [95] Yamamoto, S., "Analysis of Cesium134-Cesium137 Mixtures by Gamma-Ray Spectrometric Methods, *Analytical Chemistry*, 38, No. 9, August (1966), pp. 1261-1264.
- [96] Ying, L.-S. *et Levine*, S. P., "Fourier Transform Infrared Least-Squares Methods For the Quantitative Analysis of Multicomponent Mixtures of Airborne Vapors of Industrial Hygiene Concern", *Analytical Chemistry*, 61, (1989), pp. 677-683.
- [97] Young, M. H. *et Singhall*, N. S., "Gamma-Ray Spectral Unfolding By the Stepwise Statistical", *Nuclear Instruments and Methods*, 45, (1966), pp. 287-292.

- [98] Zhigunov, V. P., "Improvement of Resolution Function as an Inverse Problem", Nuclear Instruments and Methods 216, (1983), pp. 183-190.

- [99] Zhigunov, V. P., Kostkina, T. B. *et* Spiridov, A. A., "On Estimating Distributions with the Maximum Entropy Principle", Nuclear Instruments and Methods, A237, (1988), pp. 362-370.

Capítulo 3

LA DESCOMPOSICION ESPECTRAL VIA MEMORIA ASOCIATIVA

Analizar un espectro de radiación consiste en identificar los radionúclidos que lo componen, así como determinar las contribuciones relativas de estas formas elementales a la mezcla. En este capítulo se plantea el problema formalmente en el contexto de la teoría de regresión lineal múltiple y se demuestra que admite una solución elegante en términos de matrices pseudoinversas y, como consecuencia de ello, de memorias asociativas lineales. Se define, como resultado, un operador de este tipo que proporciona la solución al problema de la descomposición. Se propone asimismo un par de algoritmos adaptativos que construyen este operador citado.

3.1 DEFINICION DEL PROBLEMA

La revisión hecha en el capítulo anterior acerca de los problemas planteados en espectroscopía de baja resolución y el objetivo de ofrecer soluciones algorítmicas susceptibles de ser realizadas en procesadores VLSI de propósito específico ([16]) nos ha llevado a elegir las técnicas de regresión multilineal como las idóneas a estos propósitos. Siguiendo esta orientación formal, se define el problema en los siguientes términos:

- Un espectro será representado mediante un vector N -dimensional $x = [x_0, \dots, x_{N-1}]^T$ donde N es el número de canales de información del espectro y x_n es el número de eventos detectados en un determinado intervalo de energía para una fuente dada, con una instrumentación y unas condiciones experimentales (temperatura, etc.) concretas. Obviamente, x_n será un entero con valor no negativo:

$$x_n \geq 0, \quad 0 \leq n \leq N-1 \quad (3.1)$$

- Un espectro de referencia es un espectro producido por una fuente compuesta por un sólo radionúclido. Denotaremos un espectro de referencia como r_k , con $1 \leq k \leq K$. El conjunto de K vectores de referencia será denominado conjunto de referencia, que puede representarse como:

$$R = [r_1, \dots, r_K] \quad (3.2)$$

- El conjunto de espectros compuestos es el conjunto de todos los espectros que pueden ser reproducidos por combinaciones del conjunto de referencia. Un espectro compuesto está sujeto en su formación a las siguientes suposiciones:

- 1) **Principio de proporcionalidad.** Dados dos espectros, $x_m(T_1)$ y $x_m(T_2)$, obtenidos de la misma fuente emisora, usando la misma instrumentación y bajo las mismas condiciones experimentales, durante dos intervalos de tiempo

diferentes T_1 y T_2 , estos diferirán sólo en sus respectivas amplitudes, comparadas canal a canal, de acuerdo a un factor multiplicativo, que será proporcional a la razón entre los respectivos intervalos de tiempo empleados en la formación de ambos espectros:

$$\frac{x_{mn}(T_1)}{x_{mn}(T_2)} = \frac{T_1}{T_2}; \quad 0 \leq n \leq N-1 \quad (3.3)$$

- 2) **Principio de superposición:** Un espectro dado, producido por una fuente emisora compuesta por una combinación de fuentes diferentes contenidas en el conjunto de referencia, tomado en las mismas condiciones experimentales, con la misma instrumentación que los elementos del conjunto de referencia, puede ser considerado como una combinación lineal del conjunto de referencia:

$$x = Rc = \sum_{i=1}^K c_i r_i \quad (3.4)$$

donde $c = [c_1, \dots, c_K]^T$ es un vector de números reales nulos o positivos que representan los pesos o proporciones con los que cada espectro de referencia participa en la formación del espectro compuesto.

- 3) **El ruido oscuro del detector.** La respuesta del equipo de medida, en ausencia de emisión radiactiva, es un espectro característico llamado espectro de fondo y que va a ser añadido al conjunto de referencia. Los dos principios anteriores se aplican al mismo, ya que las cuentas acumuladas en cada canal crecen proporcionalmente al tiempo y, por otra parte, todo espectro recogido en el analizador multicanal puede desdoblarse en el espectro de fondo correspondiente al tiempo de conteo, más el hipotético espectro tomado en ausencia de fondo.
- 4) **Fluctuaciones estadísticas.** Como se expuso en el primer capítulo, los procesos de recolección de carga en el detector y el ruido en la electrónica de tratamiento de señal introducen invariablemente un componente errático en el espectro

obtenido en el analizador multicanal. Esta incertidumbre estadística se traduce en la adición de un número aleatorio de cuentas e_i sobre cada canal i , formando un vector e de dimensión N que se superpone al vector teórico fruto de la combinación matemática de la ecuación (3.4). Consideraremos e como el ruido que distorsiona al vector x :

$$x = Rc + e \quad (3.5)$$

Teniendo en cuenta las anteriores definiciones e hipótesis, el problema de la descomposición y cuantificación de espectros de radiación se define como sigue:

"Dado un espectro compuesto x , de dimensión N , asumiendo que todos los espectros elementales que lo forman son conocidos y están incluidos en una matriz R de dimensiones $N \times K$, queremos evaluar la estimación óptima del vector de pesos c , de dimensión K , que operado con R resulte el espectro compuesto Rc que minimice la diferencia con x en algún sentido."

Vamos a formalizar el criterio de minimización anterior buscando el vector o el espacio de vectores c que hacen mínima la norma euclídea del vector de error e . Definiendo E como la norma de e , se trata de

$$\underset{(c)}{\text{Minimizar}} \quad E(c) = \|x - Rc\|^2 \quad (3.6)$$

que es un problema clásico de mínimos cuadrados.

3.2 INTRODUCCION A LA TEORIA DE PSEUDOINVERSAS

En este apartado se examina la estrecha relación existente entre el concepto de la matriz pseudoinversa R^* de una matriz R y los problemas de optimización cuadrática lineal. Además, se introduce un algoritmo genérico de cálculo de la pseudoinversa de una matriz, que será útil posteriormente debido a sus propiedades adaptativas.

3.2.1 PSEUDOINVERSA Y MINIMOS CUADRADOS

El concepto de matriz pseudoinversa es una generalización de la inversa de una matriz cuadrada y no singular para cualquier matriz, sea rectangular, sea cuadrada pero singular ([2]). El concepto original de pseudoinversa vino dada por el siguiente teorema de existencia:

Teorema. Para cualquier matriz R de dimensión $N \times K$, la matriz

$$R^* = \lim_{\delta \rightarrow 0} (R^T R + \delta^2 I)^{-1} R^T = \lim_{\delta \rightarrow 0} R^T (R^T R + \delta^2 I)^{-1}$$

siempre existe y se denomina *matriz pseudoinversa*.

Sin embargo, la caracterización de la matriz pseudoinversa a partir de los lemas de Penrose en 1955 fue el acicate de un renovado interés en el campo de las inversas generalizadas de matrices. La pseudoinversa de una matriz real R fue definida por Penrose como la única matriz R^* que satisface las siguientes condiciones:

$$R^* R R^* = R^*,$$

$$R R^* R = R,$$

$$(RR^*)^T = RR^*,$$

$$(R^*R)^T = R^*R.$$

La pseudoinversa ha sido ampliamente investigada y aplicada ([2] y sus referencias, [12]), ya que proporciona de forma compacta descripciones de operadores geométricos de proyección y caracteriza las soluciones al problema de los mínimos cuadrados:

Teorema. Para cualquier matriz R de dimensión $N \times K$, y cualquier vector x de dimensión N , $c = R^*x$ es el vector de mínima norma de entre aquellos que minimizan $E(c) = \|x - Rc\|^2$.

El teorema anterior proporciona la solución al problema de mínimos cuadrados de una forma explícita a través de la aplicación al vector x del operador matricial R^* .

Una orientación diferente al problema de los mínimos cuadrados conduce a la solución no desde la perspectiva más elegante y directa de la matriz pseudoinversa, sino a través de la solución de las ecuaciones normales por métodos clásicos ([26]): La función escalar K -dimensional no negativa

$$E(c) = \|x - Rc\|^2$$

alcanza el mínimo en aquellos puntos en los que se anulan todas sus derivadas parciales:

$$\frac{\delta E}{\delta c_i} = 0, \quad 1 \leq i \leq K.$$

Este conjunto de ecuaciones lineales deviene en

$$R^T R c = R^T x, \tag{3.7}$$

universalmente conocidas como las "ecuaciones normales". Bajo este enfoque, para cada

espectro problema x debe ser resuelto un nuevo conjunto de ecuaciones lineales. Frente a las ecuaciones normales, la aproximación del operador pseudoinversa es ventajosa por su simplicidad, tanto conceptual como operativa, ya que puede caracterizarse todo el espacio de soluciones de las ecuaciones normales en términos de pseudoinversa como demuestra el siguiente teorema.

Teorema. c_0 minimiza $\|x - Rc\|^2$ si y sólo si c_0 es de la forma

$c_0 = R^+x + (I - R^+R)y$ para todo vector N -dimensional y . El valor de c que minimiza c_0 es único si y sólo si $R^+R = I$. Esto es cierto si y sólo si las columnas de R son vectores linealmente independientes.

Como puede apreciarse, un caso especial del problema (3.6) es aquel en el que los vectores columna de R , $\{r_1, \dots, r_K\}$, son linealmente independientes. Entonces la matriz R^TR tiene el rango completo K , con lo que su inversa está definida y la solución a las ecuaciones normales es única y viene dada por $c = (R^TR)^{-1}Rx$. En la teoría de pseudoinversas, la condición de independencia se traduce en el teorema siguiente.

Teorema. $R^+ = (R^TR)^{-1}R^T$ si las columnas de R son linealmente independientes.

Este resultado permite obtener la solución al problema (3.6) de dos formas diferentes. La primera es la clásica y es la utilizada en toda la bibliografía consultada en lo que se refiere al análisis multivariado en espectroscopía gamma ([8]) y química analítica ([14]), confiando en técnicas tradicionales de resolución de sistemas lineales. La segunda ([33]) es la que adoptaremos en este trabajo.

Una de las propiedades más interesantes de la pseudoinversa es que se comporta como un operador de proyección. Más adelante utilizaremos esta propiedad geométrica de la matriz pseudoinversa como criterio de rechazo de un espectro problema al abordar su descomposición por no ajustarse al modelo descrito su definición. La Fig. 3.1 ilustra el siguiente teorema:

Teorema. Sean los vectores $\{r_1, \dots, r_k\}$ las columnas de la matriz R . Entonces, para cualquier x , $\hat{x} = RR^+x$ es la proyección ortogonal de x sobre el subespacio expandido por $\{r_1, \dots, r_k\}$.

A continuación se introduce un algoritmo de construcción de pseudoinversas con propiedades adaptativas que lo hace interesante en la construcción de memorias asociativas lineales.

3.2.2 EL ALGORITMO DE GREVILLE

Existen varios métodos de cómputo de la pseudoinversa de una matriz ([2], [39], [45], [47]). De entre todos ellos, el algoritmo de Greville resulta interesante debido a que las operaciones que conlleva son básicamente productos matriz-vector. Ello propicia su realización sobre procesadores VLSI de producto interno ([16]). Además, esta peculiaridad la hace muy útil de cara a una implementación en tecnología óptica. Por otra parte, la independencia lineal de las referencias tratadas conducen a una simplificación del algoritmo.

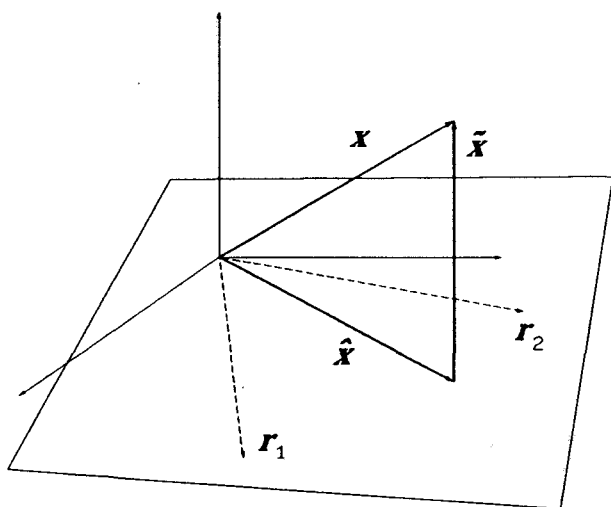


Fig. 3.1 El operador RR^+ proyecta un vector x sobre el subespacio expandido por las columnas de la matriz $R = [r_1, r_2]$.

Presentamos a continuación el algoritmo de Greville ([2], [25]). La idea básica de este algoritmo es particionar la matriz original en columnas y reclutarlas una tras otra. Descrito "grosso modo", cada vez que tomamos una nueva columna de la matriz original, computamos una nueva fila de la matriz pseudoinversa y recomputamos la submatriz complementaria.

Teorema. Sea la matriz $R_k = [R_{k-1} | r_k]$, siendo R_{k-1} la submatriz que contiene las $k-1$ primeras columnas y r_k la última de ellas. Entonces:

$$R_k^* = \begin{bmatrix} R_{k-1}^*(I - r_k p_k^T) \\ \hline p_k^T \end{bmatrix} \quad (3.8)$$

donde

$$p_k = \begin{cases} \frac{(I - R_{k-1} R_{k-1}^*) r_k}{\| (I - R_{k-1} R_{k-1}^*) r_k \|^2} & \text{si el numerador es } \neq 0, \\ \frac{(R_{k-1}^*)^T R_{k-1}^* r_k}{1 + \| R_{k-1}^* r_k \|^2} & \text{en otro caso.} \end{cases} \quad (3.9)$$

El valor inicial de R_1 es la primera columna de la matriz R por lo que $R_1^* = r_1^T (r_1^T r_1)^{-1}$, si r_1 no es el vector nulo y $R_1^* = 0^T$ si r_1 es el vector nulo. Con los datos iniciales de R_1 y R_1^* , el algoritmo de Greville no es más que la aplicación en cada iteración el teorema del mismo nombre.

3.3 MEMORIAS ASOCIATIVAS LINEALES

Una memoria asociativa, en su acepción más general, puede definirse como un operador o función de transferencia que hace corresponder un conjunto de vectores o patrones de entrada con un conjunto de vectores o patrones de salida. Si la correspondencia es robusta, un vector ruidoso cercano (respecto a alguna métrica) a un patrón de entrada debe ser asignado a un vector cercano al patrón de salida correspondiente. Los modelos de redes neurales ([3], [23]) más populares como el perceptrón multicapa ([28], [44], [51]) y redes tipo Hopfield ([4], [18], [19], [49]) son operadores no lineales que realizan transformaciones asociativas de un modo robusto. Sin embargo, no resultan apropiados en el problema de la descomposición espectral. Si bien estas estructuras matemáticas se comportan de forma adecuada en problemas de reconocimiento de patrones ([27], [50]), son incapaces de tratar correctamente la cuantificación de componentes en el problema de la mezcla, debido precisamente a la discrepancia existente entre su carácter fuertemente *no lineal* y el hecho de que el modelo de formación de un espectro de radiación sea un proceso fundamentalmente *lineal*. La descomposición de un espectro mezcla es un problema de regresión, es decir, de optimización cuadrática, más que de clasificación y como tal debe ser abordado.

Los trabajos en los campos de las redes neurales tipo Hopfield y circuitos de optimización han sido unificados y puestos bajo una perspectiva global por Kennedy y Chua ([20]-[22]). Estos autores han aplicado herramientas analíticas propias de la teoría de circuitos eléctricos no lineales a la red de Hopfield, lo cual ha permitido entender ésta como un caso particular de sistema dinámico que, para una determinada inyección de corrientes o patrón de estímulo, evoluciona hacia una configuración estacionaria de tensiones y corrientes que minimiza la energía interna del sistema, definida esta como su *función de co-contenido* ([20], [21]). A partir de ese momento comienzan a aparecer en la bibliografía aplicaciones concretas de estas redes en problemas inversos ([9], [31], [35]). Una buena revisión del estado actual de las redes neurales en tareas de optimización cuadrática puede encontrarse en [7].

Capítulo esencial es la realización física de este tipo de dispositivos. En 1984, Chua

y Lin ([6], [53]) introdujeron un circuito basado en amplificadores operacionales y fuentes de corriente no lineales capaz de solucionar cualquier problema de optimización no lineal. El problema de la mezcla, tal y como se ha planteado, es un problema de optimización lineal, por lo cual puede acometerse utilizando alguna realización de este tipo de dispositivos. El atractivo de los mismos reside en su capacidad de llegar a la solución del problema en tiempo real, lo que les hace interesantes en aplicaciones críticas de control, robótica, etc. Las desventajas son fundamentalmente dos. La primera es su rigidez. La adición de una nueva variable en el problema requiere un rediseño de todo el circuito. La segunda es su falta de exactitud. Los amplificadores operacionales están sometidos a derivas en ganancia y 'offset' debido al envejecimiento y, sobre todo, a cambios en la temperatura. Las fuentes no lineales acusan problemas de estabilidad y falta de precisión. Las mejoras en la tecnología base de estos elementos, como la introducción de los condensadores conmutados ([43]) o CCD's ([37]) palía este problema, pero no lo elimina. Según datos de la referencia [6], problemas sencillos de tres variables proporcionaban errores relativos del orden del tres por ciento. Este dato no puede ser dado por suficientemente bueno en una aplicación de análisis radiactivo, ya que da lugar a la falsa detección de radionúclidos en un espectro mezcla que realmente no los contiene. En el caso sumamente realista de que el espectro mezcla venga afectado por derivas instrumentales, la adición de esta considerable fuente de error en el tratamiento matemático de los datos descarta por completo una aproximación analógica al problema de la mezcla, al menos con la tecnología de hoy día.

Ante el aún incierto panorama que ofrece la solución analógica, en este trabajo se adopta la siguiente aproximación al problema: Se opta por la formación adaptativa de una memoria lineal a través un algoritmo apropiado ([10]), susceptible de ser implementado sobre estructuras aritméticas digitales de alta escala de integración ([1], [16], [42]). Con ella se gana, por una parte, flexibilidad en la incorporación de referencias y, por otra, precisión en el análisis respecto a una solución analógica. Se pierde en rapidez. Sin embargo, este aspecto no resulta crítico en la aplicación tratada, como se discutirá posteriormente.

3.3.1 LA MEMORIA LINEAL OPTIMA

Este formalismo se fundamenta en la existencia de transformaciones lineales entre patrones espaciales que pueden ser diseñadas *óptimamente* en el sentido de robustez ante estímulos ruidosos ([5], [32], [48]). Los resultados matemáticos coinciden básicamente con los de teoría de estimación y regresión multilínea ([36]), pudiendo unificarse todos ellos en el mismo formalismo de la teoría de pseudoinversas. El problema se plantea en los siguientes términos. Sean los patrones de entrada $x_k \in R^N$ donde $1 \leq k \leq K$, y sean los patrones de salida correspondientes $y_k \in R^M$. Se trata de encontrar el operador matricial M que lleve a cabo esta función de una forma óptima, minimizando el error sobre todos los y_k . Agrupando los patrones como columnas de las matrices X e Y , el problema se formula como:

$$\underset{(M)}{\text{Minimizar}} \quad \|Y - MX\|^2 \quad (3.10)$$

Esta ecuación admite en principio todo un hiperplano de soluciones. La teoría de pseudoinversas establece que la solución de mínima norma es $M = YX^*$ a través del siguiente teorema ([2], [25]):

Teorema. Si X e Y son matrices con el mismo número de columnas, entonces las soluciones al problema de minimización de la función matricial $E(M) = \|Y - MX\|^2$ vienen dadas en la forma $M = YX^* + Z[I - XX^*]$ para cualquier Z de las mismas dimensiones que M , siendo $M = YX^*$ la de mínima norma. La solución $M = YX^*$ es única si y sólo si las columnas de X son linealmente independientes.

3.3.2 COMPUTACION RECURSIVA DE MEMORIAS LINEALES

La teoría desarrollada en la formación adaptativa de memorias asociativas ([24], [25], [32]), basadas en trabajo previo sobre teoría de pseudoinversas ([2], [39], [45]) y proceso adaptativo de señal ([30], [52]), constituye una solución computacional alternativa a los métodos clásicos de resolución de sistemas lineales ligados a rutinas estándar de mínimos cuadrados ([26]). Considerar el operador solución a las ecuaciones normales $(R^T R)^{-1} R$ como una memoria supone, citando a Kohonen ([25]), "construirlo a través de un *proceso adaptativo* mediante la influencia progresiva de los vectores de referencia". Es más, esta vía puede considerarse como un avance conceptual al introducir los conceptos de *estructuras de cómputo y algoritmos de adaptación* de las mismas. Este paradigma, gestado en la teoría de filtrado adaptativo y redes neurales artificiales, lleva a la realización física de las estructuras y al desarrollo de algoritmos de adaptación de las mismas en tareas específicas ([18], [42]). En este contexto, vamos a examinar la algoritmia asociada a la formación adaptativa de memorias lineales.

Sea M la memoria lineal asociativa óptima que transforma los patrones de entrada x_i en patrones de salida y_i mediante la transformación lineal $y_i = Mx_i$, $1 \leq i \leq k-1$. El problema que trata esta sección es el siguiente: si sabemos que M es óptima para los pares (x_i, y_i) , $1 \leq i \leq k-1$, entonces se trata de averiguar cuál es la *corrección* que hay que efectuar sobre M para que su nuevo valor sea óptimo sobre los pares (x_k, y_k) , donde $1 \leq i \leq k$. Este es un problema al que denominamos *recursivo* ya que el nuevo valor óptimo de M_k es una función del valor óptimo previo M_{k-1} y de las nuevas observaciones (x_k, y_k) . Todos los procesos recursivos en este contexto obedecen al esquema:

$$M_k = M_{k-1} + (y_k - M_{k-1}x_k)c_k^T \quad (3.11)$$

donde c_k^T define la corrección que debe efectuarse sobre el *error de predicción* $y_k - M_{k-1}x_k$

de la memoria anterior M_{k-1} sobre el nuevo par (x_k, y_k) . Distintos algoritmos de adaptación progresiva de la memoria se derivan de distintas formas de escoger c_k^T . A continuación se introducen dos de ellos, el algoritmo de proyección del gradiente y el algoritmo de Widrow-Hoff.

3.3.2.1 EL ALGORITMO DE PROYECCION DEL GRADIENTE

La ecuación (3.11) se concreta mediante un vector de corrección basado en el algoritmo de Greville. Si en el paso k $M_k = Y_k X_k^*$, aplicando (3.8) tenemos que

$$\begin{aligned}
 M_k &= Y_k X_k^* = [Y_{k-1} | y_k] \left[\begin{array}{c} X_{k-1}^*(I - x_k p_k^T) \\ \hline p_k^T \end{array} \right] \\
 &= Y_{k-1} X_{k-1}^* + (y_k - Y_{k-1} X_{k-1}^* x_k) p_k^T \\
 &= M_{k-1} + (y_k - M_{k-1} x_k) p_k^T,
 \end{aligned} \tag{3.12}$$

que determina en un sólo paso la configuración de una nueva memoria M_k que asocia óptimamente los pares (x_i, y_i) , $1 \leq i \leq k$, en función éstos y de la memoria anterior M_{k-1} .

La ecuación anterior puede simplificarse si tenemos en cuenta el siguiente argumento geométrico: la matriz $X_{k-1} X_{k-1}^*$, de dimensión $N \times N$, es un operador que proyecta el vector x_k sobre el subespacio expandido por los vectores x_1, \dots, x_{k-1} , resultando el vector \hat{x}_k . De este modo, podemos siempre descomponer x_k en la forma $x_k = \hat{x}_k + \tilde{x}_k$ tal y como ilustra la Fig. 3.1. Si todos los vectores de referencia son linealmente independientes, el vector $\tilde{x}_k = \Phi_{k-1}(x_k) \equiv (I - X_{k-1} X_{k-1}^*) x_k$ es no nulo y la opción superior de (3.9) es la que siempre se toma como vector p_k . En estas condiciones,

$$p_k^T = \frac{\tilde{x}_k^T}{\|\tilde{x}_k\|^2} \quad (3.13)$$

y el algoritmo recursivo presenta el aspecto final que sigue:

$$M_k = M_{k-1} + (y_k - M_{k-1}x_k) \frac{\tilde{x}_k^T}{\|\tilde{x}_k\|^2}$$

El cómputo del operador $\Phi_k \equiv I - X_k X_k^T$ admite la siguiente solución recursiva:

$$\Phi_k = \Phi_{k-1} - \frac{\tilde{x}_k \tilde{x}_k^T}{\|\tilde{x}_k\|^2}$$

con $\tilde{x}_k = \Phi_{k-1}(x_k)$ y las condiciones iniciales de $\Phi_0 = I$ y $\tilde{x}_1 = x_1$. Sin embargo, este método de cálculo supone una carga computacional alta al tener que trabajar en cada paso con matrices de orden $N \times N$. Una vía alternativa es usar el procedimiento de ortogonalización de Gram-Schmidt:

$$\tilde{x}_i = x_i - \sum_{j=1}^{i-1} \frac{(x_i^T \tilde{x}_j) \tilde{x}_j}{\|\tilde{x}_j\|^2} \quad (3.14)$$

Este procedimiento fue utilizado por Pyle ([39]) en resolución de sistemas lineales, y en la literatura posterior se le ha dado en llamar el *método de proyección del gradiente*.

3.3.2.2 EL ALGORITMO DE WIDROW-HOFF

El problema planteado en (3.10) puede acometerse dando una solución inicial M_0 y tratando de refinarla en iteraciones sucesivas buscando nuevas configuraciones M_i que lleven a un valor $E(M_i) = \|Y - M_i X\|^2$ cada vez menor hasta llegar a una solución óptima.

El conjunto de tales soluciones viene dado por $M = YX^* + Z[I - XX^*]$ y un procedimiento de este tipo no garantiza la de mínima norma porque depende de la aproximación inicial. Sin embargo, en el caso de independencia lineal de las columnas de X , sabemos que la solución es única y por lo tanto independiente de la configuración inicial. Uno de los métodos de aproximación sucesiva a una solución M que minimice $E(M)$ es el método del gradiente. El gradiente de $E(M)$ viene dado por la expresión $G_E(M) = 2(MX - Y)X^T$. Se trata, entonces, de avanzar en el espacio de configuraciones en la dirección del gradiente en cada punto, pero en sentido opuesto, en pasos o incrementos que vienen de dados por un parámetro λ , que puede ser fijo o depender de la iteración. Así, podemos dar la siguiente regla de avance para minimizar $E(M)$:

$$\begin{aligned} M_0 & \text{ arbitraria} \\ M_i &= M_{i-1} - \lambda_i(M_{i-1}X - Y)X^T \end{aligned} \tag{3.15}$$

Observar que el c_k de la ecuación (3.11) vale en este caso $\lambda_k X$. Si la secuencia $\{\lambda_i\}$ cumple que $\sum_{k=1}^{\infty} \lambda_k = +\infty$ y $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$ entonces el proceso converge a una de las memorias óptimas caracterizadas por $M = YX^* + Z[I - XX^*]$.

El problema de hallar una memoria óptima puede verse como una clase más general de problemas que es el de la *estimación de parámetros lineales*. Su formulación trata de encontrar una relación lineal entre los pares (x_k, y_k) , $1 \leq k \leq K$, donde x_k es un vector de dimensión N de variables observadas que tratamos de relacionar linealmente con y_k , otro vector de variables observadas de dimensión M , de tal modo que ([30]):

$$y_k = P^T x_k + v_k. \tag{3.16}$$

v_k es el vector de error de la regresión. Por fin, P^T , de dimensión $M \times N$, es la matriz de parámetros que tratamos de estimar a partir de los K vectores de observación. En forma

matricial, (3.16) queda:

$$Y = P^T X + V \quad (3.17)$$

donde no es difícil darse cuenta de que la minimización de la norma cuadrática del residuo V conduce a la misma estimación paramétrica que el de una hipotética memoria asociativa entre las observaciones X e Y en los términos anteriormente discutidos.

En aplicaciones de control adaptativo ([11]), (3.16) se utiliza para determinar parámetros en ecuaciones de diferencias. En comunicaciones y proceso adaptativo de señal ([15], [52]) es el modelo estándar para ecualizadores, filtros de reducción de ruido, cancelación de eco, etc.

Es importante hacer notar que podemos adaptar cada una de las M filas de P^T por separado y de modo paralelo. Así entendido, el proceso de construcción de todas y cada una de las filas de la matriz de parámetros se transforma en M procesos de adaptación de una fila p_m^T a fin de que:

$$Y_{mk} = p_m^T x_k + V_{mk} \quad 1 \leq k \leq K. \quad (3.18)$$

Cada una de estas filas puede ser entendida como un filtro lineal de orden N , con lo cual el problema de estimar una memoria o función de correspondencia óptima entre pares de observaciones (x_k, y_k) puede ser abordado en el área estimación paramétrica lineal, predicción lineal, filtros digitales e identificación de sistemas discretos ([15], [41], [52]).

Más específicamente, en el contexto de adaptación de filtros lineales, podemos citar tres algoritmos fundamentales. El primero es el conocido como "steepest descent" o descenso más pronunciado. Realiza un descenso en la función de error siguiendo la dirección del gradiente del mismo modo que (3.15). El segundo es el RLS ([30], [52]) o de mínimos cuadrados recursivo, más rápido pero más complejo, por que utiliza informa-

ción de correlación entre los regresores P^T para llevar a cabo el descenso en la función de error de un modo más eficiente que el anterior. Por último, uno de los algoritmos más estudiados y utilizados en la práctica es el de el *gradiente estocástico*, también conocido como de Widrow-Hoff en la literatura de señal, así como la *regla delta* en teoría de redes neurales y sistemas adaptativos ([28]). Es básicamente un algoritmo de gradiente, pero en cada paso la corrección no avanza en la dirección óptima del gradiente negativo, sino en otra subóptima pero de computación más ágil. Así, volviendo al ámbito de memoria asociativa, mientras que en (3.15) cada iteración necesita del conocimiento completo del par de matrices (X, Y) para efectuar la corrección en la dirección del gradiente, la regla delta sólo exige el conocimiento de un par (x_k, y_k) de los patrones de entrada y salida almacenados en (X, Y) . Todos estos pares se introducen de forma secuencial y repetidamente disminuyendo el error de la función objetivo en cada patrón hasta que se logra la convergencia a una configuración estable o punto fijo M que, por supuesto, pertenece al conjunto $M = YX^* + Z[I - XX^*]$. Si denotamos la secuencia cíclica de "aprendizaje" $(x_1, y_1), (x_2, y_2), \dots, (x_K, y_K), (x_1, y_1), (x_2, y_2), \dots$ como $(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m), (x^{m+1}, y^{m+1}), \dots$, la regla de descenso puede escribirse como:

$$M_0 \quad \text{arbitraria} \quad (3.19)$$

$$M_i = M_{i-1} - \lambda_i (M_{i-1} x^{i+1} - y^{i+1}) (x^{i+1})^T$$

Las condiciones de convergencia del algoritmo se encuentran en las referencias [30] y [52]. Para añadir un patrón (x_{K+1}, y_{K+1}) a la memoria basta comenzar con la configuración de memoria que es ya operativa y comenzar de nuevo el proceso iterativo anterior, esta vez incluyendo el nuevo par en la secuencia. Presumiblemente, la matriz inicial no estará lejos de la que buscamos, con lo cual la readaptación llevará poco tiempo o, al menos, menos que el que llevaría una adaptación que progresase desde una configuración inicial arbitraria completamente ajena a las características comunes nuestros espectros de referencia (no negatividad, normas similares, estructura de picos sobre un fondo suave, etc).

3.4 DESCOMPOSICION ESPECTRAL VIA MEMORIA ASOCIATIVA

En el apartado anterior hemos examinado los aspectos más relevantes del campo de memoria asociativa lineal, tanto teóricos como computacionales. A continuación se deriva la memoria apropiada al problema del reconocimiento de espectros de radiación y los procedimientos específicos para construirla.

3.4.1 DERIVACION DE LA MEMORIA

Un caso especial en el contexto de memoria asociativa es aquel en el que se trata de *identificar* el patrón de entrada a la memoria. Cambiemos la notación de los patrones de entrada x_1, \dots, x_K por r_1, \dots, r_K y definamos los patrones de salida en la forma de vectores K -dimensionales $y_1 = [1, 0, \dots, 0]^T$, $y_2 = [0, 1, \dots, 0]^T$, ..., $y_K = [0, 0, \dots, 1]^T$, con lo que $Y = [y_1, \dots, y_K]$ resulta ser la matriz identidad. En el supuesto de independencia lineal de los patrones de entrada agrupados en R , la memoria viene dada por el operador pseudoinversa R^* .

En virtud de la linealidad de este operador, puede atribuírsele el principio de superposición, a través del cual al aplicar a la memoria $M = R^*$ una entrada que sea una combinación lineal de las columnas de R , la salida obtenida es el vector de coeficientes de la combinación. Así la entrada $2r_i$ provoca la salida $y = [0, \dots, 2, \dots, 0]^T$ con el componente no nulo en la posición i -ésima, que puede interpretarse como que estamos excitando la memoria con el patrón de referencia r_i , pero con intensidad doble al de referencia utilizada en la construcción de $M = R^*$. Más generalmente, la excitación de la combinación $x = Rc$, donde Rc es un vector de coeficientes reales de dimensión K , provoca una respuesta:

$$y = Mx = R^* Rc = Ic = c$$

Una importante conclusión es que la memoria lineal óptima que realiza la correspondencia entre el par (R, I) es el mismo operador que proporciona la solución al problema de descomposición espectral planteado en (3.6). A efectos operativos, este resolutor puede ser implementado en alguna tecnología, sea electrónica integrada ([1], [42]) u óptica, de modo que la descomposición puede ser llevada a cabo en tiempo real. A este respecto es obligado admitir que la velocidad en el análisis de contribuciones no es en general un requisito crítico en el área de la espectroscopía gamma ya que, para fuentes de baja actividad, el tomar su espectro puede llevar minutos. Sin embargo, sí es importante en aplicaciones recientes de espectroscopía de imágenes ([29]) donde tenemos un espectro por pixel y, por lo tanto, los tiempos de análisis se multiplican de forma cuadrática. En esta línea de trabajo vamos a reelaborar la algoritmia genérica de formación adaptativa de memorias lineales presentada en el anterior apartado acomodando los algoritmos de Pyle y Widrow-Hoff a la casuística del problema de la descomposición espectral.

3.4.2 ESPECIFICACION DE ALGORITMOS

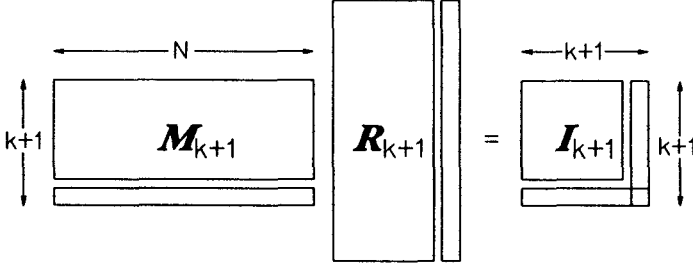
Con independencia del algoritmo de adaptación utilizado, una vez construida una memoria óptima M_k para k espectros de referencia $\{r_1, \dots, r_k\}$, esta tiene k filas y N columnas, siendo N el número de canales de los espectros. La matriz de objetivos es una matriz identidad de k filas y k columnas, cada una de estas últimas representando el código de identificación del espectro de referencia correspondiente según:

$$\begin{array}{c} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \begin{array}{c} \xrightarrow{N} \\ \hline \end{array} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \end{array} \begin{array}{c} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \begin{array}{c} \xleftarrow{k} \\ \hline \end{array} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \end{array} = \begin{array}{c} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \begin{array}{c} \xleftarrow{k} \\ \hline \end{array} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \end{array}$$

$M_k \quad R_k \quad I_k$

La memoria pasa tener $k+1$ filas cuando se introduce un nuevo espectro de referencia. El número de vectores de identificación se incrementa en uno, el correspondiente a la

nueva referencia, este último con los k primeros componentes nulos y el $(k+1)$ -ésimo la unidad, es decir, $[z_k \mid 1]^T$, y al resto se le añade un componente nulo, pasando a agruparse todos ellos en una matriz identidad de $k+1$ filas por $k+1$ columnas:



Se puede expresar la matriz de las k primeras filas de M_{k+1} como M_{k+1}^* y la fila $(k+1)$ -ésima como m_{k+1} , por lo que la nueva memoria se configura como:

$$M_{k+1} = \begin{bmatrix} M_{k+1}^* \\ \hline m_{k+1} \end{bmatrix} \quad (3.20)$$

El problema de la formación de M_{k+1} se reparte en dos tareas. Primera, la adaptación de M_{k+1}^* para acomodar el nuevo par (r_{k+1}, z_k) y, segunda, la formación de m_{k+1} con el requisito de que $m_{k+1} R_{k+1} = [z_k \mid 1]^T$. Nos ocupamos ahora de la formulación de los algoritmos de Pyle y Widrow-Hoff para este caso de estudio.

3.4.2.1 EL ALGORITMO DE PROYECCION DEL GRADIENTE

Dado que los espectros de referencia son algebraicamente linealmente independientes y que la matriz de objetivos Y_k en el paso k es la matriz identidad I_k de k filas por k columnas, el cálculo de la matriz M_k tal que $M_k R_k = I_k$ se reduce a calcular la pseudoinversa $M_k = R_k^*$. A medida que vamos incorporando nuevos espectros de referencia, nuevas

pseudoinversas R_{k+1}^* , R_{k+2}^* , etc, deben ser computadas, por lo que la aplicación directa de la reformulación de Pyle al algoritmo de Greville es una aproximación apropiada. La condición de independencia lineal permite identificar (3.20) con (3.8), con lo cual

$$m_{k+1} = \frac{\tilde{r}_{k+1}^T}{\|\tilde{r}_{k+1}\|^2}$$

donde

$$\tilde{r}_i = r_i - \sum_{j=1}^{i-1} \frac{(r_i^T \tilde{r}_j) \tilde{r}_j}{\|\tilde{r}_j\|^2}$$

y

$$M_{k+1}^* = M_k^* (I - r_{k+1} m_{k+1}^T)$$

Ante estas premisas, podemos precisar el siguiente algoritmo de adaptación de una memoria asociativa para reconocimiento de espectros de radiación:

Paso 1:

- $\tilde{r}_1 = r_1$
- $m_1 = \frac{\tilde{r}_1^T}{\|\tilde{r}_1\|^2}$
- $M_1 = m_1$

Paso k:

- $\tilde{r}_k = r_k - \sum_{j=1}^{k-1} \frac{(r_k^T \tilde{r}_j) \tilde{r}_j}{\|\tilde{r}_j\|^2}$
- $M_k^* = M_{k-1} - (M_{k-1} r_k) \frac{\tilde{r}_k^T}{\|\tilde{r}_k\|^2}$

$$-m_k = \frac{\tilde{r}_k^T}{\|\tilde{r}_k\|^2}.$$

3.4.2.2 EL ALGORITMO DE WIDROW-HOFF

Supongamos que disponemos de k espectros de referencia. El algoritmo de Widrow-Hoff calcula una memoria asociativa óptima M_k a partir de una configuración arbitraria inicial $M_k^{(0)}$ mediante un descenso progresivo y convergente en la función de error pasando por estados intermedios sucesivos $M_k^{(i)}$ hasta alcanzar un criterio dado de terminación. En el caso que nos ocupa, al introducir el espectro de referencia $k+1$, la matriz M_{k+1} aumenta en una fila según la expresión (3.20). Es necesario entonces, por una parte, fijar los componentes o pesos de m_{k+1} y, por otra, readaptar los de M_k para obtener la matriz M_{k+1}^* . En cuanto a la primera tarea, podemos partir de una fila o filtro N -dimensional nulo $m_{k+1}^{(0)}$ y tratar de adaptarlo según el criterio clásico de minimización del error cuadrático, de modo que configuraciones sucesivas $m_{k+1}^{(i)}$ converjan a m_{k+1} . En cuanto a la segunda, una configuración inicial de M_{k+1}^* que a priori se presenta como una buena aproximación es $M_{k+1}^{*(0)} = M_k$. Sucesivas configuraciones $M_{k+1}^{*(i)}$ nos llevarán al estado final M_{k+1}^* . Podemos particionar la nueva matriz de objetivos I_{k+1} de forma análoga a la de la memoria M_{k+1} como sigue:

$$I_{k+1} = \begin{bmatrix} I_{k+1}^* \\ \text{-----} \\ v_{k+1} \end{bmatrix}. \quad (3.21)$$

Y ya que z_k es un vector columna nulo k -dimensional, podemos expresar a su vez I_{k+1}^* como $I_{k+1}^* = [I_k \mid z_k] = [u_1, \dots, u_k, u_{k+1}]$. De este modo, M_{k+1}^* es el operador óptimo que lleva

a cabo la transformación entre los pares (r_i, u_i) , $1 \leq i \leq k+1$. Podemos escribir la $(k+1)$ -ésima fila de I_{k+1} , v_{k+1} , como $v_{k+1} = [z_k^T \mid 1]$. Así, se puede entender m_{k+1} como el filtro óptimo que lleva a cabo el ajuste entre la secuencia vectorial $\{r_1, \dots, r_{k+1}\}$ y la secuencia escalar $\{z_k^T, 1\}$. Esta última puede escribirse como $\{v_i\}$, $1 \leq i \leq k+1$. Teniendo presente la notación introducida, se propone el siguiente procedimiento:

$$\begin{aligned} \text{Paso } 1: \quad & - m_1^{(0)} = 0 \\ & - m_1^{(j+1)} = m_1^{(j)} - \lambda^{(j)}(m_1^{(j)} r^{(j)} - 1) r^{(j)T} \\ & - M_1 = m_1 \end{aligned}$$

$$\begin{aligned} \text{Paso } k: \quad & - M_k^{*(0)} = M_{k-1} \\ & - m_k^{(0)} = 0 \end{aligned}$$

Desde $j = 1$ hasta un valor de terminación dado:

$$\begin{aligned} & - M_k^{*(j)} = M_k^{*(j-1)} - \lambda^{(j)}(M_k^{*(j-1)} r^{(j)} - u^{(j)}) r^{(j)T} \\ & - m_k^{(j)} = m_k^{(j-1)} - \lambda^{(j)}(m_k^{(j-1)} r^{(j)} - v^{(j)}) r^{(j)T} \end{aligned}$$

3.5 IMPLEMENTACION VLSI DE LA MEMORIA ASOCIATIVA

Una descripción detallada de todos aquellos aspectos relativos a la integración en silicio de la memoria asociativa excede con mucho los límites de este trabajo. No obstante, sí consideramos de interés el destacar algunas ideas generales, dado que la integración de la algoritmia constituye una de sus líneas maestras ([10], [42]). Dos aproximaciones son posibles en función del grado de flexibilidad que se pretenda otorgar a la memoria. La primera consiste en integrar una memoria fija con pesos precalculados. No importa el

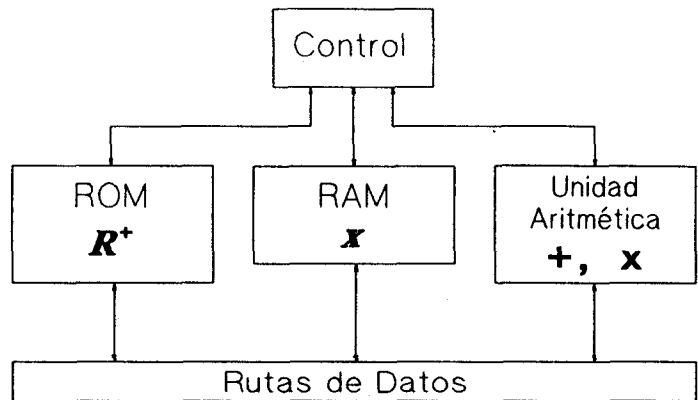


Fig. 3.4 Implementación de una memoria no adaptativa que evalúa $c = R^+ x$

algoritmo de cómputo, ni si es adaptativo o no lo es, puesto que éste se realiza "off-line". La memoria R^+ se almacena en memoria ROM y el vector problema x en memoria RAM (Fig. 3.4). El vector de contribuciones se obtiene a través de la operación $c = R^+ x$, que implica K productos internos, por lo que la unidad aritmética del sistema únicamente necesita de las operaciones de suma y multiplicación. La rigidez de esta solución es evidente pero caso de resultar suficiente, la rapidez de la descomposición está asegurada. Una segunda aproximación, mucho más ambiciosa, es aquella que persigue la integración de los algoritmos de proyección del gradiente o de Widrow-Hoff, que dotan de adaptatividad a la memoria. La misión de los mismos consiste en construir "on line" la memoria que

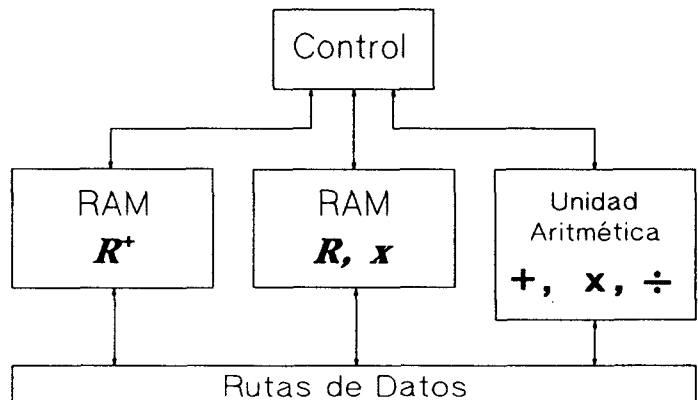


Fig. 3.5 Implementación de una memoria adaptativa que evalúa $c = R^+ x$ y del algoritmo de Pyle que la construye.

descompone una mezcla problema de forma óptima, incorporando a la misma las referencias necesarias y suficientes de entre todas las disponibles en el conjunto. Entraremos con posterioridad en estos detalles. Las diferencias más interesantes entre ambos algoritmos residen en sus respectivas cargas computacionales y en sus requerimientos en aritmética. En la referencia [10] se desarrollan con más detalle estos aspectos. Por una parte, el algoritmo de Pyle presenta una complejidad computacional más bien determinista que depende del orden K del resolutor M_K , pero de cara a la integración en silicio requiere de la implementación de la operación de división (Fig. 3.5). Por su parte, el algoritmo de Widrow-Hoff requiere sólo las operaciones de suma y multiplicación (Fig. 3.6), si bien el criterio de terminación de las iteraciones depende de factores de convergencia ligados a la correlación de los vectores del conjunto de referencia y al máximo error admitido para dar el operador como bueno. A causa de ello, no es posible conocer a priori el número de iteraciones que demanda un problema dado, pero sí es posible derivar una conclusión importante de esta discusión acerca de la implementación de memorias asociativas, y es que el diseño de la arquitectura y la aritmética de un resolutor de este tipo debe ser siempre precedido de un estudio algebraico detallado.

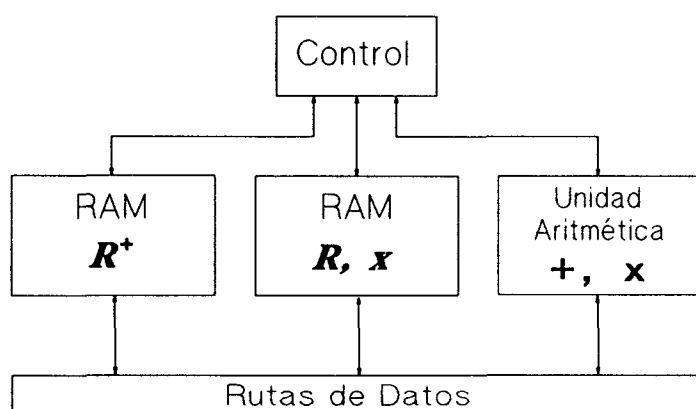


Fig. 3.6 Implementación de una memoria que evalúa $c=R^+x$ y del algoritmo de Widrow-Hoff que la construye.

3.6 CRITERIOS DE ESTIMACION DEL ERROR

El método de descomposición de un espectro problema x a través de una memoria lineal óptima proporciona un vector c de contribuciones relativas de los elementos de referencia almacenados como columnas de R . Por supuesto, la estimación de c nunca es exacta, debido a la naturaleza estadística de la emisión radiactiva y la acción del ruido en la instrumentación. Incluso puede llegar a ser completamente disparatada si el espectro problema no responde al modelo de formación lineal descrito en el primer apartado. Existen dos factores fundamentales que provocan la divergencia de un espectro de radiación respecto al modelo de formación:

- 1) La participación en la mezcla de espectros correspondientes a radionúclidos no contemplados en el conjunto de referencia.
- 2) La aparición de derivas en la instrumentación por efecto de las condiciones experimentales en las que se captura el espectro problema.

Si bien la distorsión provocada en la mezcla por las derivas puede compensarse en ciertas condiciones mediante un procedimiento corrector como veremos en el siguiente capítulo, el carecer de un conjunto de referencia suficiente para explicar la mezcla es un problema insalvable y que da lugar a contribuciones erróneas ([33], [46]). Aunque algunas publicaciones recientes que aplican implementaciones optoelectrónicas de redes tipo Hopfield al problema de la mezcla ([4], [29]) ignoran este hecho, es absolutamente necesario, al menos, detectar la insuficiencia del modelo cuando se realiza un análisis de contribuciones. En tal caso, los resultados no pueden siquiera ser considerados como aproximaciones, sino simplemente erróneos. A partir de la interpretación geométrica del problema resulta sencillo establecer un criterio para rechazar de forma categórica un análisis de contribuciones. En el supuesto de que el análisis sea validado, consideraciones estadísticas proporcionan un intervalo de confianza de cada contribución estimada.

3.6.1 VALIDACION DEL MODELO DE FORMACION

En la Fig. 3.7 el vector problema x se descompone en la suma de dos vectores ortogonales, la proyección \hat{x} sobre el espacio expandido por el conjunto de referencia y la componente ortogonal no proyectable \tilde{x} . Vamos a llamar α al ángulo que forman x y \hat{x} según la Fig. 3.7. Si la mezcla problema puede ser enteramente resuelta como combinación lineal de los espectros del conjunto de referencia, la componente \tilde{x} será nula y el ángulo α también. En otro caso, aparece una componente no proyectable \tilde{x} . Es intuitivo pensar que cuanto mayor sea α mayor es la desviación respecto del modelo de formación. Incluso si el modelo es válido, el ruido que siempre se superpone al vector problema provoca un α no nulo. Resulta inmediato, entonces, discriminar entre un espectro que se ajusta al modelo y otro que no a partir de un ángulo umbral α_u . Si $\alpha \leq \alpha_u$ el análisis de contribuciones se acepta. De lo contrario se rechaza.

3.6.2 ESTIMACION DEL ERROR DEL VECTOR DE CONTRIBUCIONES

El propósito de este apartado es el de establecer un nivel de confianza en la estimación de cada uno de los componentes del vector de contribuciones c una vez adoptado el

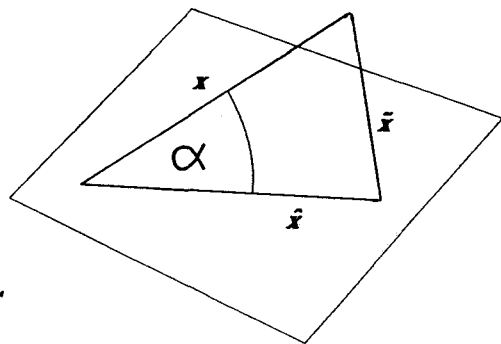


Fig. 3.7 Representación geométrica de la discrepancia entre la mezcla problema y el modelo de formación.

método de descomposición de la memoria lineal. Para ello es necesario revisar algunos métodos estadísticos de estimación como son el Modelo de Regresión Lineal Múltiple y el Método de Máxima Verosimilitud, vistos en el capítulo anterior aplicados ambos al problema de la mezcla. El método de máxima verosimilitud estima las contribuciones como el vector \hat{c} solución al problema:

$$\underset{(c)}{\text{Minimizar}} (x-Rc)^T \Sigma_x^{-1}(x-Rc) \quad (3.22)$$

donde Σ_x es la matriz de covarianza dependiente del espectro problema x . En realidad esta dependencia existe y es más bien compleja [8]. Una simplificación tradicional es el método de los mínimos cuadrados ponderados como vimos en el capítulo anterior. Puede simplificarse aún más la estimación, sin pérdida sensible de precisión ([33], [34]), haciendo $\Sigma_x = \sigma^2 I_N$. El problema (3.22) se simplifica en el siguiente:

$$\underset{(c)}{\text{Minimizar}} (x-Rc)^T(x-Rc) \quad (3.23)$$

que conduce a las ecuaciones normales $(R^T R)c = R^T x$ cuya solución es $\hat{c} = (R^T R)^{-1} R^T x$. Así, el estimador de máxima verosimilitud del vector de contribuciones es:

$$\hat{c} = R^+ x \quad (3.24)$$

ecuación que pone de manifiesto que el hallar una memoria asociativa óptima en el sentido de minimizar el error cuadrático entre las asociaciones, y la estimación paramétrica de máxima verosimilitud arriba planteada son problemas equivalentes. Esta analogía servirá para estimar el error del vector de contribuciones \hat{c} proporcionado por la memoria.

Es bien conocido en regresión lineal ([8]) que el estimador \hat{c} de máxima verosimilitud proporcionado por (3.24) tiene una distribución normal de media c y matriz de covarianzas $Cov(\hat{c}) = \sigma^2 (R^T R)^{-1}$ y que un estimador centrado de σ^2 es

$$\sigma^2 = \frac{\tilde{\mathbf{x}}^T \tilde{\mathbf{x}}}{N - K - 1} . \quad (3.25)$$

Llamando q_{ij} a los elementos de $(R^T R)^{-1}$ tenemos que $\text{Var}(\hat{c}_j) = \sigma^2 q_{jj}$ por lo que un intervalo de confianza del 95% para \hat{c} es:

$$[\hat{c}_j - 2\sigma\sqrt{q_{jj}} < c_j < \hat{c}_j + 2\sigma\sqrt{q_{jj}}] \quad (3.26)$$

El valor de $\sqrt{q_{jj}}$ es conocido a priori, pues depende sólo de la matriz de espectros de referencia R . El valor de σ^2 depende de cada análisis en particular y se calcula según (3.25). Esto hace que para dar un margen de error sea necesario hallar el vector de residuo $\tilde{\mathbf{x}} = \mathbf{x} - R\mathbf{c}$.

El error cuadrático medio de \hat{c} respecto del valor real viene dado por ([36]):

$$\begin{aligned} ECM(\hat{c}) &= \sum_{i=1}^K (\hat{c}_i - c_i)^2 \\ &= \sigma^2 \text{tr}[(R^T R)^{-1}] \\ &= \sigma^2 \sum_{i=1}^K \frac{1}{\lambda_i} \end{aligned} \quad (3.27)$$

siendo tr la traza de una matriz cuadrada o suma algebraica de los elementos de la diagonal principal y siendo λ_i los autovalores de la matriz de correlación $R^T R$. Cuanto más parecidos son los espectros de referencia o bien cuanto mayor es el número de ellos, es decir, menor es su distancia angular, más cerca estará $R^T R$ de la singularidad, lo que significa que el autovalor mínimo más pequeño de los λ_i (todos son positivos por ser $R^T R$ definida positiva ([13])) se acerca a cero, aumentando el valor del error cuadrático de la estimación de contribuciones según (3.27). Como consecuencia, puede decirse que la calidad del análisis viene condicionado por la ortogonalidad de los espectros de referencia.

En el siguiente apartado se propone un criterio de elección y eliminación de referencias en el proceso de construcción de la memoria donde aplicaremos las consideraciones aquí expuestas.

Por otra parte, observaciones experimentales han puesto en evidencia que la norma de la memoria R^* aumenta con la colinealidad o la semejanza entre los espectros de referencia contenidos en R . Esta semejanza disminuye, logicamente, al aumentar el poder de resolución espectral de la instrumentación. El interés en llevar a cabo una investigación sobre métodos de análisis de espectros de baja o muy baja resolución hace necesario proporcionar una justificación teórica este efecto de amplificación de la norma. En primer lugar, tomemos como definición de norma de una matriz la norma de Frobenius, también llamada norma euclídea. Así, la norma de una matriz R se define como $\| R \|_F \equiv (tr[R^T R])^{1/2}$. Ya que $R^T R$ es definida positiva, todos sus autovalores son positivos. Sean estos $\lambda_1 \geq \dots \geq \lambda_k$. Y debido a que la traza de una matriz coincide con la suma algebraica de sus autovalores, tenemos que $\| R \|_F = (\lambda_1 + \dots + \lambda_k)^{1/2}$. Sabemos, por otra parte, a través del teorema de la descomposición singular ([13]), que los autovalores de $(R^*)^T(R^*)$ son

$$\frac{1}{\lambda_k} \geq \frac{1}{\lambda_{k-1}} \geq \dots \geq \frac{1}{\lambda_1},$$

por lo que

$$\| R^* \|_F = \left(\frac{1}{\lambda_k} + \dots + \frac{1}{\lambda_1} \right)^{1/2}.$$

A medida que un aumento en la colinealidad en R provoca un acercamiento de $R^T R$ a la singularidad, el autovalor λ_k se acerca a cero. Así, y debido a la positividad de todos los λ_i :

$$\lim_{\lambda_r \rightarrow 0} \| R^* \| = \lim_{\lambda_r \rightarrow 0} \left(\frac{1}{\lambda_K} + \dots + \frac{1}{\lambda_1} \right)^{1/2} = \infty$$

que justifica el efecto de amplificación observado. Este hecho ha de tenerse en cuenta a la hora de una implementación digital de la memoria, ya que para una precisión en bits determinada provocará una pérdida de resolución debido al aumento el rango dinámico de los componentes de R^* .

3.7 BUSCANDO EL OPERADOR MINIMO

Cuando se aborda un problema de regresión múltiple pueden darse dos situaciones posibles. En la primera, el número de funciones de ajuste es conocido a priori y únicamente es necesario establecer los coeficientes óptimos. Este el caso típico del ajuste polinomial de una función en un dominio de definición dado. En el segundo caso, los regresores no son conocidos a priori, sino que pertenecen a un conjunto muy amplio y es necesario escoger aquellos que proporcionan un resultado óptimo teniendo en cuenta una serie de restricciones. La descomposición de un espectro mezcla responde a esta última situación, en la que a priori el número de candidatos a participar en la mezcla es alto, a saber, el de todos los isótopos posibles. Afortunadamente, el valor de la contribución de un isótopo a la mezcla es una cantidad no negativa. Este hecho introduce una restricción importante a la hora de decidir la inclusión de los mismos en el conjunto que explica la mezcla. Así, este conjunto queda sensiblemente reducido.

Ante este complejo panorama, se plantea el problema de encontrar la memoria asociativa construida con el número de referencias necesario y suficiente (operador mínimo) que descompone una mezcla problema o conjunto de ellas de la misma composición en lo que a isótopos participantes se refiere. Ya que la solución no es conocida a priori, se impone un método iterativo de refinamiento progresivo de la base de referencias consideradas. Emergen dos estrategias opuestas en la búsqueda de candidatos. La primera consiste en

realizar un ajuste con el conjunto de referencia completo, eliminar candidatos de contribuciones negativas o casi nulas e iniciar el proceso de nuevo. Como es fácil apreciar, este enfoque no es apropiado de cara una implementación algorítmica en procesadores a medida, ya que el tamaño del problema es demasiado grande en sus primeros pasos y esto es precisamente lo que se trata de evitar. La alternativa es la técnica opuesta, consistente en ir incorporando referencias a la memoria de una forma adaptativa como se ha sugerido en apartados previos. En cada paso de iteración, se incorporará a la memoria el candidato que presente mayor similitud angular con la mezcla problema. La mezcla se presentará de nuevo a la memoria, esta vez ampliada, y el análisis de contribuciones determinará la parada del proceso por haber llegado a la configuración idónea y por tanto a una evaluación final correcta, o es preciso una nueva iteración.

Dos situaciones determinan la necesidad de una nueva iteración tras presentar la mezcla problema a la memoria:

- 1) Son detectadas referencias que provocan contribuciones negativas. Estas deben ser eliminadas del conjunto de referencia actual y construir de nuevo la memoria.
- 2) El ángulo α es demasiado alto y por lo tanto se estima necesario incorporar un nuevo candidato para explicar la mezcla.

Este tipo técnicas y otras derivadas de ellas pertenecen al campo de la llamada "regresión por pasos", de la que el lector encontrará cumplida información en [17] y [36], por nombrar algunas. La orientación a una implementación VLSI de un algoritmo de selección de candidatos lleva a la decisión de formación progresiva de la memoria como ya hemos apuntado. A continuación se precisan estos principios en el siguiente algoritmo de selección de candidatos:

```

 $K = \{\phi\};$ 
 $e := x;$ 
DO
{
     $\rho_i = \cos^2(r_i, e);$ 
    Sea  $g / \rho_g = \max_i \rho_i;$ 
     $K := K \cup \{g\};$ 
DO
{
     $R := [\cdots | r_i | \cdots], i \in K;$ 
     $M := R^*;$ 
     $c := Mx;$ 
     $N := \{i / c_i \leq NivelSig\};$ 
     $K := K - \{N\};$ 
} WHILE ( $N \neq \{\phi\}$ );
 $e := [I - RM]x;$ 
 $sen^2(\alpha) := e^T e / x^T x;$ 
} WHILE ( $sen^2(\alpha) < UMBRAL$ );
return ( $M$ );

```

(3.28)

El algoritmo dispone de un universo de referencias r_i , $i \in Q$ y es alimentado por la mezcla problema x . Del conjunto Q selecciona el subconjunto $K \subset Q$ que contiene las referencias necesarias y suficientes para explicar la mezcla problema x y a partir de este construye la memoria mínima M que la descompone. El conjunto K está inicialmente vacío y es el bucle externo el encargado de incorporar nuevos candidatos al mismo. A su vez, el bucle interno se encarga de eliminar de K aquellos candidatos anteriormente escogidos tales que tras aumentar K con una nueva referencia y construir la memoria correspondiente provocan ahora contribuciones negativas o bien por debajo de un nivel de significancia dado que se considera inferior a la propia incerteza de la mezcla. La fórmula del intervalo de confianza del 95 % (3.26) resulta orientadora a este respecto, si bien difícilmente aplicable por el requerimiento de cómputo de la raíz cuadrada que complicaría enormemente el diseño

del procesador. Un umbral convenido parece la aproximación más razonable. La terminación del proceso queda determinada cuando el grado de ajuste α de la Fig. 3.7 se considera el apropiado.

El $\cos^2(r_i, e)$ es una *medida de similitud* angular entre la referencia candidata y el vector de error. El cuadrado del coseno resulta más conveniente porque evita las raíces cuadradas propias de la definición de la norma euclídea. Así, tenemos que:

$$\cos^2(r_i, e) = \frac{(r_i^T e)^2}{r_i^T r_i e^T e} \quad (3.29)$$

La utilidad de este criterio de similitud que se propone estriba en que sólomente se requiere una única operación de división para computarlo, lo que no supone una carga adicional excesiva de cálculo de cara a una implementación completa VLSI del algoritmo que no disponga de la operación división en su unidad aritmético-lógica ([10], [42]). El mismo argumento puede utilizarse con respecto a la medida global de encaje:

$$\text{sen}^2(\alpha) = \frac{e^T e}{x^T x} \quad (3.30)$$

que también requiere una única división.

El algoritmo (3.28) ha sido programado en el entorno PC386/MS-DOS. Su comportamiento ha sido siempre correcto, llegando a descomponer mezclas que tomaban sus componentes de un conjunto de hasta treinta candidatos de 512 canales, que fue el límite que impuso el sistema operativo MS-DOS debido a su baja capacidad de direccionamiento de 640K ([38]).

3.8 BIBLIOGRAFIA

- [1] Aguayo, P. *et* Gómez, P., "Memoria asociativa para reconocimiento de espectros", VI Congreso de Diseño de Circuitos Integrados, Universidad de Cantabria, Santander (España), 12-15 Noviembre, (1991), pp. 125-130.
- [2] Albert, A., "Regression and the Moore-Penrose Pseudoinverse", Academic Press, (1972).
- [3] Anderson, J. E., *et* Rosenfeld, A., Eds., "Neurocomputing: Foundations Of Research", MIT Press, (1988).
- [4] Barnard, B. *et* Casasent, D. P., "Optical Neural Net For Classifying Imaging Spectrometer Data", Applied Optics, Vol. 28, No. 15, August, (1989), pp. 3129-3133.
- [5] Cherkassky, V. *et al.*, "Linear Algebra Approach to Neural Associative Memories and Noise Performance of Neural Classifiers", IEEE Trans. on Computers, Vol. 40, No. 12, December, (1991), pp. 1429-1435.
- [6] Chua, L. O. *et* Lin, G. N., "Nonlinear Programming without Computation", IEEE Trans. on Circuits and Systems, Vol. CAS-31, No. 2, February, (1984), pp. 182-188.
- [7] Cichocki, A. *et* Umbehauen, R., "Neural Networks for Solving Systems of Linear Equations and Related Problems", IEEE trans. on Circuits and Systems, Vol. CAS-39, No. 2, February, (1992), pp. 124-138.
- [8] Covell, D. F. *et al.*, "Multiple Linear Regression Analisis of Scintillation Gamma-Ray Spectra: Theoretical and Practical Considerations", Nuclear Instruments and Methods, 80, (1970), pp. 55-60.
- [9] Culhane, A. D. *et al.*, "A Neural Net Approach to Discrete Hartley and Fourier Transformations", IEEE Trans. on Circuits and Systems, Vol. 36, No. 5, May, (1989), pp. 695-703.
- [10] Díaz, J. C. *et al.*, "An Associative Memory to Solve the Mixture Problem in Composite Spectra", 35th Midwest Symposium on Circuits and Systems, Washington, August 9-12, (1992), pp. 891-894.

- [11] Fu, K. S., "Learning Control Systems, Review and Outlook", IEEE Transactions on P.A.M.I., Vol PAMI-8, No. 3, May, (1986), pp. 327-342.
- [12] Greville, T. N. E., "Some Applications of the Pseudoinverse of a Matrix", SIAM review, Vol. 2, No. 1, (1960), pp. 15-21.
- [13] Griffel, D. H., "Linear Algebra and its Applications", John Wiley & Sons, (1989).
- [14] Haaland, D. M., "Quantitative Infrared Analysis of Borphosphosilicate Films Using Multivariate Statistical Methods", Analytical Chemistry, 60, (1988), pp. 1208-1217.
- [15] Haykin, S., "Introduction to Adaptive Filters", MacMillan Publishing Company. (1984).
- [16] Higgins, R., "Digital Signal Processing in VLSI", Ed. Prentice-Hall, (1990).
- [17] Hogan, M. A. *et al.*, "Multiple Linear Regression Analysis of Scintillation Gamma-Ray Spectra: Automatic Candidate Selection", Nuclear Instruments and Methods, 80, (1970), pp. 61-68.
- [18] Hopfield, J. J., "Neurons with Graded Response Have Collective Computatinal Properties Like Those of Two-State Neurons" Proc. Natl. Acad. Sci. USA, Vol. 81, (1984), pp. 3088-3092.
- [19] Hopfield, J. J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Soc. Natl. Acad. Sci. USA, Vol. 79, April, (1982), pp. 2554-2558.
- [20] Kennedy, M. P. *et* Chua, L. O., "Unifying the Tank and Hopfield Linear Programming Circuit and the Canonical Nonlinear Programming Circuit of Chua and Lin", IEEE Trans. on Circuits and Systems, Vol. CAS-34, No. 2, February, (1987), pp. 210-214.
- [21] Kennedy, M. P. *et* Chua, L. O., "Circuit Theoretic Solutions for Neural Networks - An Old Approach to a New Problem", Proceedings of the IEEE First International Conference on Neural Networks, Vol. 2, San Diego, June 21-24, (1987), pp. 169-176.
- [22] Kennedy, M. P. *et* Chua, L. O., "Neural Networks for Nonlinear Programming", IEEE Trans on Circuits and Systems, Vol. 35, No. 5, May, (1988), pp. 554-562.

- [23] Khanna, T., "Foundations of Neural Networks", Addison-Wesley, (1990).
- [24] Kohonen, T. *et* Oja, E., "Fast Adaptive Formation of Orthogonalizing Filters and Associative Memory in Recurrent Networks of Neuron-Like Elements", *Biological Cybernetics*, 21, (1976), pp. 85-95.
- [25] Kohonen, T., "Self-Organization and Associative Memory", Second Edition, Springer-Verlag, (1988).
- [26] Lawson, C. L. *et* Hanson, R. J., "Solving Least Squares Problems", Prentice Hall, (1974).
- [27] Liddell, R. W. *et* Jurs, P. C., "Interpretation of Infrared Spectra Using Pattern Recognition Techniques", *Analytical Chemistry*, Vol. 46, No. 14, December, (1974), pp. 2126-2132.
- [28] Lippmann, R. P., "An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, Vol. 3, No. 4, (1987), pp. 4-22.
- [29] Liu, S.D. *et* Casasent, D. P., "Optical Processing of Imaging Spectrometer Data", *SPIE*, Vol. 938 Digital and Optical Shape Representation and Pattern Recognition, (1988), pp.212-220.
- [30] Ljung, L., "Analysis of Stochastic Gradient Algorithms for Linear Regression Problems", *IEEE Trans. on Information Theory*, Vol. 30, No. 2, (1984), pp. 151-160.
- [31] Marrian, M. C. *et* Peckerar, C. R. K., "Electronic 'Neural' Net Algorithm for Maximum Entropy Deconvolution", *Proceedings of the IEEE First International Conference on Neural Networks*, San Diego, June 21-24, (1987).
- [32] Murakami, K. *et* Aibara, T., "An Improvement On the Moore-Penrose Generalized Inverse Associative Memory", *IEEE Trans. on Systems, Man and Cybernetics*, Vol. SMC-17, No. 4, July/August, (1987), pp. 699-707.
- [33] Olmos, P., Díaz, J. C. *et al.*, "A New Approach To Automatic Radiation Spectrum Analysis", *IEEE Trans. on Nuclear Science*, Vol. 38, No. 4, August, (1991), pp. 971-975.
- [34] Olmos, P., Díaz, J. C. *et al.*, "Application of Neural Network Techniques In Gamma Spectroscopy", *Nuclear Instruments and Methods*, A312, (1992), pp. 167-173.

- [35] Pati, Y. C. *et al.* "Neural Networks for Tactile Perceptions", Proc. IEEE Conf. on Robotics and Automation", Philadelphia, (1988), pp. 136-139.
- [36] Peña, D., "Estadística: Modelos y Métodos. 2: Modelos Lineales y Series Temporales". Alianza Editorial, (1989).
- [37] Peterson, C., "Track Finding with Neural Networks", Nuclear Instruments and Methods, Vol. A279, (1989), pp. 537-545.
- [38] Prosise, J., "DOS-5.0 Memory Management", Ziff-Davis Press, (1992).
- [39] Pyle, L. D., "Generalized Inverse Computations Using the Gradient Projection Method", Journal of the ACM, Vol. 11, No. 4, October, (1964), pp. 422-428.
- [40] Rao, C. R., "Optimization. Theory and Applications", Wiley Eastern Ltd. (1978).
- [41] Robinson *et al.* Silvia, "Digital Foundations Of Time Series Analysis: The Box-Jenkins Approach", Holden-Day, Inc., (1979).
- [42] Rodellar, V. *et al.*, "A VLSI Arithmetic Unit for a Signal Processing Neural Neural Network", 35th Midwest Symposium on Circuits and Systems, Washington, August 9-12, (1992), pp. 1044-1047.
- [43] Rodríguez Vázquez, A. *et al.*, "Nonlinear Switched-Capacitor 'Neural' Networks for Optimization Problems", IEEE Trans. on Circuits and Systems, Vol. 37, March, (1990), pp. 384-398.
- [44] Rumelhart, D. E. *et al.* McClelland, J. L., "Parallel Distributed Processing", MIT Press, (1988).
- [45] Rust, B. *et al.*, "A Simple Algorithm For Computing the Generalized Inverse Of a Matrix", Communications of the ACM, Vol. 9, No. 5, May, (1966), pp. 381-387.
- [46] Schonfeld, E. *et al.*, "Determination of Nuclide Concentrations in Solutions Containing Low Levels of Radiativity by Least-Squares Resolution of the Gamma-Ray Spectra", Nuclear Instruments and Methods, 45, (1966), pp. 1-22.

- [47] Soderstrom, T. *et* Steward, G. W., "On the Numerical Properties of an Iterative Method for Computing the Moore-Penrose Generalized Inverse", *SIAM J. Numer. Anal.*, Vol.11, No. 1, March, (1974), pp. 61-74.
- [48] Stiles, G. S. *et* Denq, D., "On the Effect of Noise on the Moore-Penrose Generalized Inverse Associative Memory", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 3, (1985), pp. 358-360.
- [49] Tank, D. W. *et* Hopfield, J. J., "Simple 'Neural' Optimization Networks: an A/D Converter, Signal Decision Circuit and a Linear Programming Circuit", *IEEE Trans on Circuits and Systems*, CAS-33, (1986), pp. 533-541.
- [50] Tou, J. T. *et* González, R. C., "Pattern Recognition Principles", Addison-Wesley, (1974).
- [51] Widrow, B. *et* Lehr, M. A., "30 Years of Adaptive Neural Networks: Perceptron, Madaline and Backpropagation", *Proceedings of the IEEE*, Vol. 78, No. 9, September, (1990), pp. 1415-1442.
- [52] Widrow, B. *et* Stearns, S. D., "Adaptive Signal Processing", Prentice-Hall, (1985).
- [53] Wilson, G., "Quadratic Programming Analogs", *IEEE Trans. on Circuits and Systems*, Vol. CAS-33, No. 9, September, (1986), pp. 907-911.

Capítulo 4

AMPLIACION DE LA MEMORIA ANTE PEQUEÑAS DERIVAS INSTRUMENTALES

Una memoria asociativa lineal, entendida como la pseudoinversa de la matriz de espectros del conjunto de referencia, realiza la descomposición espectral de una mezcla problema siempre que, primero, el conjunto de referencia sea suficiente para explicarla y, segundo, no existan desviaciones en la instrumentación que provoquen deformaciones en la misma. Desgraciadamente, las derivas instrumentales siempre se manifiestan en mayor o menor grado. En este capítulo se estudia el caso en el que la magnitud de la deriva puede considerarse pequeña, que es la situación más común

cuando el equipo de espectroscopía es sometido a controles periódicos. Se propone un modelo paramétrico de un espectro genérico que toma en cuenta las condiciones experimentales en las que se lleva a cabo el conteo. El desarrollo del modelo conduce a una ampliación sencilla de la memoria capaz de tolerar la deriva de la mezcla problema y llevar a cabo un análisis robusto de contribuciones.

4.1 DESCRIPCION ANALITICA DEL ESPECTRO CON DISTORSION INSTRUMENTAL

Un analizador multicanal de radiación gamma es un dispositivo que establece una correspondencia entre la magnitud de la energía e de un fotón incidente y un canal discreto x , de acuerdo con la relación lineal $x(e) = ae + b$, donde a y b son parámetros del analizador denominados respectivamente *ganancia* y *umbral* y son variables físicas independientes.

Los valores de estos parámetros pueden sufrir incrementos o decrementos por efecto de las condiciones ambientales en las que opere el sensor y la cadena de instrumentación, principalmente la temperatura ([8]). Ello provoca que una relación original energía-canal

$$x_0(e) = ae + b \tag{4.1}$$

se transforme en

$$x_1(e) = (a + \Delta a)e + (b + \Delta b) \tag{4.2}$$

La gráfica de la Fig. 4.1 ilustra las ecuaciones (4.1) y (4.2) y muestra superpuestos un espectro original y otro afectado por una fuerte deriva en ganancia y umbral a fin de

mostrar mejor su efecto. Un desarrollo de (4.2) lleva a que

$$\begin{aligned} x_1(e) &= (a+\Delta a)e+(b+\Delta b) \\ &= ae + b + \Delta ae + \Delta b \\ &= x_0(e) + \Delta ae + \Delta b \end{aligned}$$

En la configuración de calibración del analizador, podemos considerar, sin pérdida de generalidad, que $a = 1$ y $b = 0$. De esta forma:

$$x_0(e) = e \tag{4.3}$$

$$\begin{aligned} x_1(e) &= x_0(e) + \Delta a x_0(e) + \Delta b \\ &= (1 + \Delta a) x_0(e) + \Delta b \end{aligned} \tag{4.4}$$

Llamando $p \equiv 1 + \Delta a$ y $q \equiv \Delta b$, (4.4) resulta:

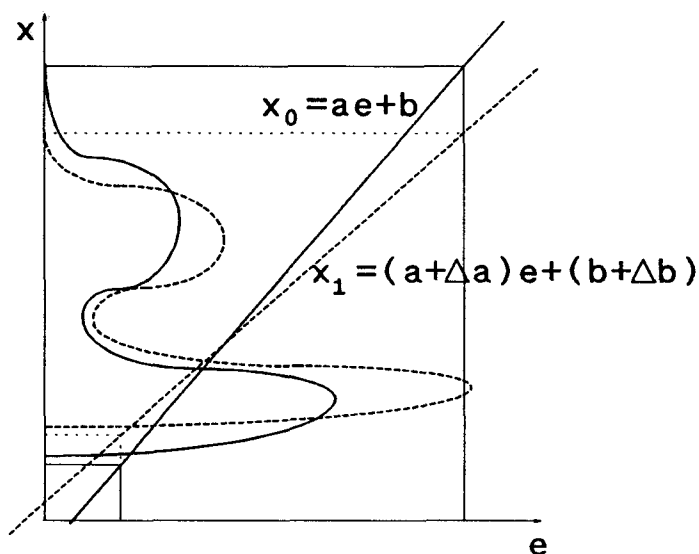


Fig. 4.1 Las derivas instrumentales provocan una nueva relación lineal energía-canal. Este cambio se traduce en una deformación del espectro, observada principalmente en un desplazamiento de los fotopicos.

$$x_1 = px_0 + q \quad (4.5)$$

que no es más que un cambio de variable lineal. Tras estas consideraciones, podemos afirmar que un espectro cualquiera $f(x)$, al sufrir una distorsión $(\Delta a, \Delta b)$ pasa a ser un espectro definido analíticamente por la nueva función $\alpha f(px+q)$. El parámetro α puede determinarse utilizando el principio de conservación del conteo ([6]), que indica que ambos espectros tienen el mismo número de cuentas. Expresado matemáticamente:

$$\int_{t_1}^{t_2} f(x) dx = \alpha \int_{\frac{t_1-q}{p}}^{\frac{t_2-q}{p}} f(px+q) dx$$

donde los límites de integración representan el rango de la variable independiente en el cual la función es mayor que cero, es decir, el intervalo de canales en los que se han acumulado una o más cuentas.

Sea $I(x) + C$ la integral indefinida de $f(x)$. Entonces $(1/p)I(px + q) + C$ es la integral indefinida de $f(px + q)$, siendo C una constante arbitraria. De esta forma:

$$I(x) \Bigg|_{t_1}^{t_2} = \alpha \left[\frac{1}{p} I(px + q) \right]_{\frac{t_1-q}{p}}^{\frac{t_2-q}{p}}$$

Sustituyendo límites de integración:

$$\alpha \frac{1}{p} I(px + q) = I(t_2) - I(t_1) ,$$

con lo que $\alpha = p$. Así, puede decirse que un espectro cualquiera $f(x)$ deviene en

$$f^{(\Delta a, \Delta b)}(x) = (1 + \Delta a)f[(1 + \Delta a)x + \Delta b] \quad (4.6)$$

cuando se produce una variación $(\Delta a, \Delta b)$ en los parámetros instrumentales de ganancia y umbral, tal y como muestra la Fig. 4.1. Los efectos de la deriva en ganancia pueden ser apreciados en la Fig. 4.2. Esta muestra un par de espectros de ^{60}Co tomados con un detector experimental de centelleo de INa construido en los laboratorios del C.I.E.M.A.T. en 1990. La captura de los espectros se llevó a cabo en dos períodos de conteo separados por un intervalo de tiempo de cinco horas. Se aprecia a simple vista que se produjo una deriva considerable en ganancia en este intervalo. El lector puede formarse una idea de la relación entre la deformación que en un espectro dado provoca la deriva en ganancia y la magnitud de la misma, Δa , observando la Fig. 4.3. Esta figura recoge las diversas formas funcionales adoptadas por un espectro sintético $f((1+\Delta a)x)$ de 1024 canales al variar el parámetro de ganancia. El espectro $f((1+\Delta a)x)$ fue construido superponiendo formas gaussianas a un fondo exponencial¹.

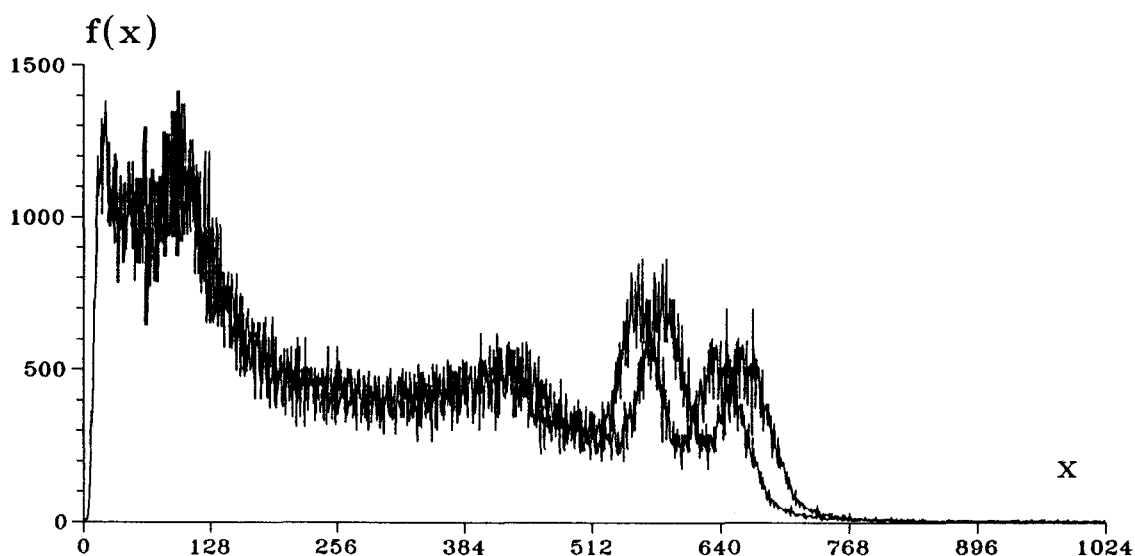


Fig. 4.2 Dos espectros de Cobalto 60 registrados bajo dos valores distintos de la ganancia del instrumento.

¹ Es necesario apuntar que el rango de variación en ganancia que presenta la figura (del -10% al +10%) resulta exagerado comparado con el esperado en la práctica y sólo cumple un propósito ilustrativo.

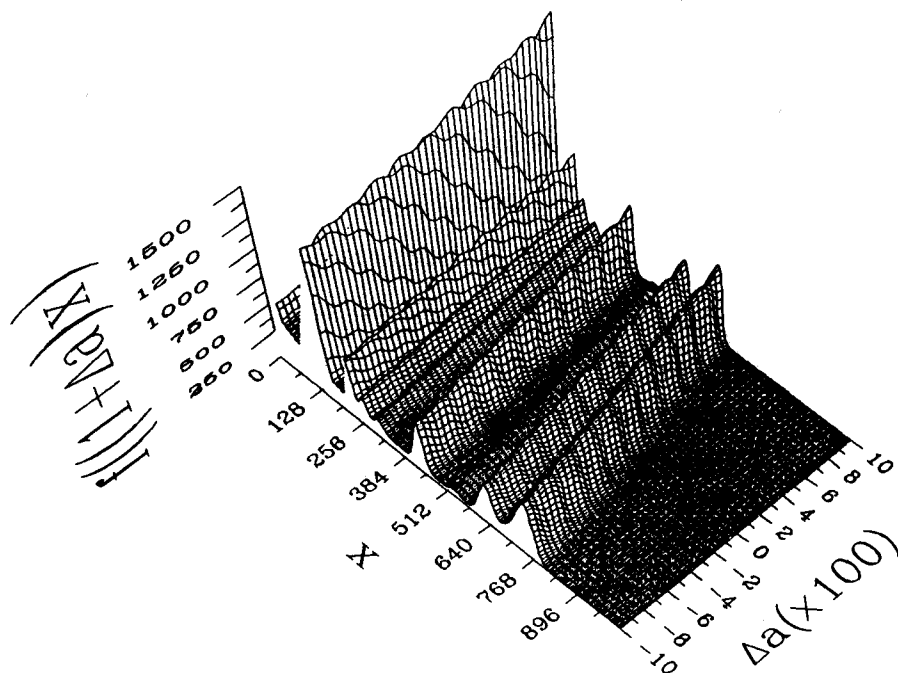


Fig. 4.3 Espectro sintético de 1024 canales en función de la deriva en ganancia

4.2 MEMORIA ASOCIATIVA Y PEQUEÑAS DERIVAS INSTRUMENTALES

Tomemos el número de canal como una variable continua y sea $f(x)$ un espectro libre de derivas. Salvo un factor de escala, el espectro correspondiente a la misma fuente, pero obtenido bajo las condiciones o parámetros de ganancia a y umbral b , puede describirse como $f(ax + b)$ según quedó establecido en el primer apartado. El teorema de Taylor, aplicado hasta el primer orden, determina el desarrollo de $f(x, a, b)$ en $a = 1$ y $b = 0$ como:

$$f(ax + b) \approx f(x) + (a - 1) \frac{df}{da} + (b - 0) \frac{df}{db} \quad (4.7)$$

La teoría de diferenciales ([1]) establece que:

$$\delta f = \frac{df}{dt} \delta t$$

por lo que

$$\frac{\delta f}{\delta a}_{(x, 1, 0)} = \frac{df}{dt} \frac{\delta t}{\delta a}_{(x, 1, 0)} = \frac{df}{dt}_{(t=x)} \frac{\delta t}{\delta a}_{(a=1)}$$

donde $t(x, a, b) = ax + b$. Así,

$$\frac{df}{da}_{(x, 1, 0)} = \frac{df}{dx} x$$

Un desarrollo similar conduce a que:

$$\frac{\delta f}{\delta b}_{(x, 1, 0)} = \frac{df}{dx}$$

La ecuación (4.7) queda entonces:

$$\begin{aligned} f(ax + b) &\approx f(x) + x \frac{df}{dx}(a - 1) + \frac{df}{dx}(b - 0) \\ &= f(x) + x f'(x) \Delta a + f'(x) \Delta b \end{aligned} \tag{4.8}$$

Teniendo en cuenta que x es una variable discreta x_n , $0 \leq n \leq N-1$, donde N es el número de canales del espectrómetro, (4.8) es expresable como una ecuación vectorial

$$\begin{bmatrix} f_0^{(\Delta a, \Delta b)} \\ \vdots \\ f_{N-1}^{(\Delta a, \Delta b)} \end{bmatrix} \approx \begin{bmatrix} f_0 \\ \vdots \\ f_{N-1} \end{bmatrix} + \Delta a \begin{bmatrix} x_0 f'_0 \\ \vdots \\ x_{N-1} f'_{N-1} \end{bmatrix} + \Delta b \begin{bmatrix} f'_0 \\ \vdots \\ f'_{N-1} \end{bmatrix}, \quad (4.9)$$

donde f_i es el número de cuentas acumuladas en el canal i en el espectro capturado sin derivas instrumentales, f'_i es el valor de la derivada del espectro respecto del canal en el canal i y $f_i^{(\Delta a, \Delta b)}$ representa el número de cuentas acumuladas en el canal i y es función de dos variables: las derivas instrumentales Δa y Δb . Basta un ejemplo para poder apreciar la forma de estos nuevos vectores. Si tomamos el espectro de cobalto de mayor ganancia de la Fig. 4.2 como el vector f_i , $0 \leq i \leq N-1$, entonces los vectores f'_i , $0 \leq i \leq N-1$, y $x_i f'_i$, $0 \leq i \leq N-1$, vienen representados en la Fig. 4.4 y la Fig. 4.5 respectivamente.

Puede darse una expresión más compacta de la ecuación (4.9) haciendo uso del concepto de la matriz jacobiana. Teniendo en cuenta que el jacobiano de un conjunto de funciones $f_i(u_1, \dots, u_M)$, donde $0 \leq i \leq N-1$, se define como

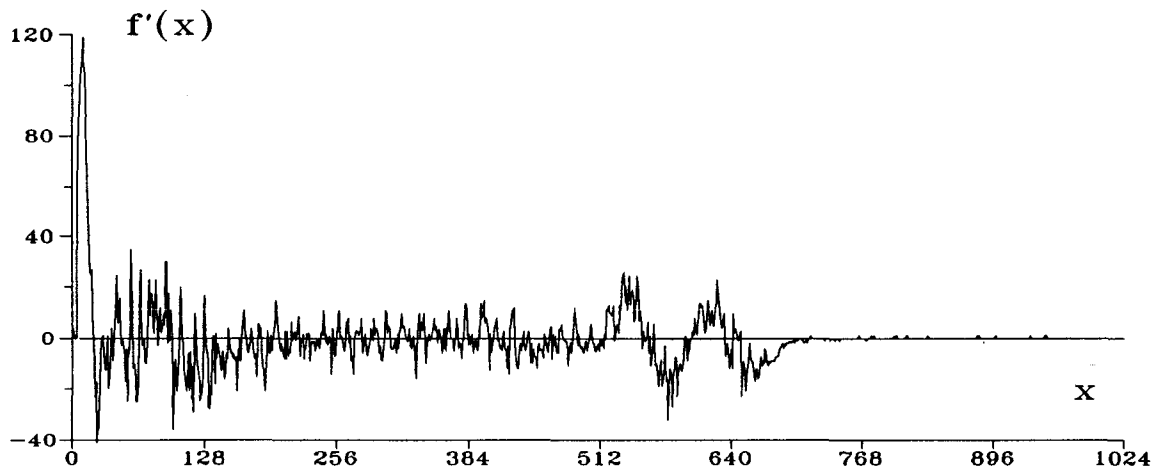


Fig. 4.4 La derivada respecto del número de canal del espectro de ^{60}Co de mayor ganancia de la Fig. 4.2.

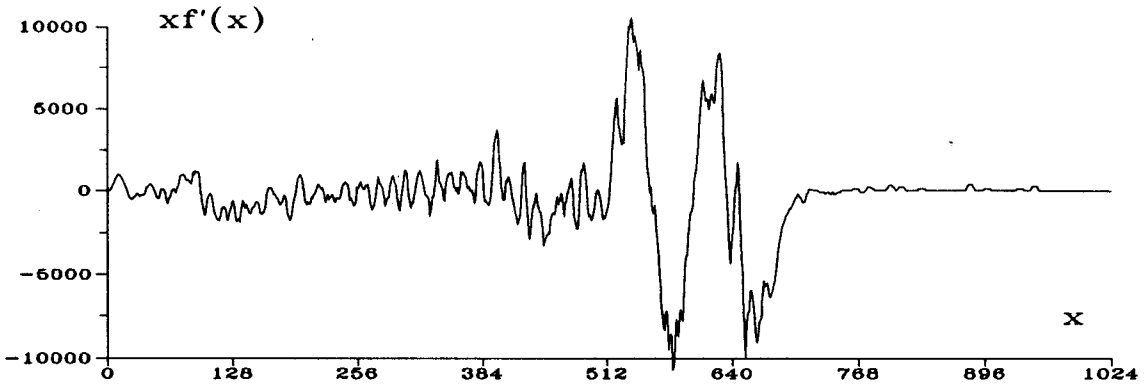


Fig. 4.5 Función derivada del espectro de ^{60}Co de mayor ganancia de la Fig. 4.2 multiplicada por el número de canal.

$$\nabla f = \begin{bmatrix} \frac{\delta f_1}{\delta u_1}, \dots, \frac{\delta f_1}{\delta u_M} \\ \vdots \\ \frac{\delta f_N}{\delta u_1}, \dots, \frac{\delta f_N}{\delta u_M} \end{bmatrix}, \quad (4.10)$$

la ecuación (4.9) puede escribirse en la forma

$$f(1+\Delta a, 0+\Delta b) = f(1, 0) + \nabla f(1, 0) [\Delta a, \Delta b]^T.$$

El interés de la ecuación radica en que es posible expresar un espectro con derivas como una *expansión lineal* cuando estas se mantienen pequeñas. Este hecho permite formalizar el problema de las derivas dentro de la teoría de las memorias asociativas lineales. La idea se desarrolla en el resto del presente capítulo y se aplica al problema de la descomposición de una mezcla afectada por derivas ([8], [11], [12]).

4.2.1 DERIVAS EN ESPECTROS PUROS

Supongamos por un momento que no existe el problema de la mezcla, es decir, sólomente disponemos de un espectro de referencia r o, en notación continua $r(x)$, y los espectros problema consisten en versiones de r afectadas por derivas y actividades diversas. La información que queremos extraer del espectro problema hipotético $g(x)$ es su actividad relativa c respecto a la del espectro de referencia $r(x)$, así como las desviaciones Δa y Δb en ganancia y umbral respectivamente. De acuerdo con el modelo paramétrico de distorsión (4.6), el espectro problema puede ser descrito como

$$\begin{aligned} g(x) &= c(1+\Delta a) r[(1+\Delta a)x + \Delta b] + e(x) \\ &= c(1+\Delta a)r(x, \Delta a, \Delta b) + e(x) \\ &= c' r(x, \Delta a, \Delta b) + e(x) , \end{aligned} \tag{4.11}$$

es decir, como una función del canal con un parámetro lineal c' , ligado a la actividad de la muestra, y dos parámetros no lineales de deriva Δa y Δb . $e(x)$ es el componente gaussiano de incertidumbre estadística que siempre acompaña al espectro.

Si las derivas son pequeñas, la aproximación de Taylor de (4.8) es correcta, y entonces $g(x)$ puede ser expandida linealmente en la forma:

$$\begin{aligned} g(x) &= c' r(x, \Delta a, \Delta b) + e(x) \\ &\approx c' [r(x) + x r'(x) \Delta a + r'(x) \Delta b] + e(x) \\ &= c' [r(x)] + c' \Delta a [x r'(x)] + c' \Delta b [r'(x)] + e(x) \end{aligned} \tag{4.12}$$

o, en notación vectorial:

$$\begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{N-1} \end{bmatrix} \approx c' \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_{N-1} \end{bmatrix} + c' \Delta a \begin{bmatrix} 0 r'_0 \\ 1 r'_1 \\ \vdots \\ (N-1) r'_{N-1} \end{bmatrix} + c' \Delta b \begin{bmatrix} r'_0 \\ r'_1 \\ \vdots \\ r'_{N-1} \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{N-1} \end{bmatrix}$$

Resumiendo, puede decirse que un espectro con una distorsión pequeña puede desdoblarse en una combinación lineal sencilla del espectro sin derivas y dos vectores fácilmente computables a partir de este: su derivada respecto del canal y esta derivada multiplicada por la abscisa o el número de canal. De esta forma, la dependencia del espectro problema es *lineal* respecto del vector de componentes c' , $c' \Delta a$ y $c' \Delta b$. Si convenimos en denominar

$$r^{(0)} = [r_0, \dots, r_{N-1}]^T,$$

$$r^{(1)} = [r'_0, \dots, r'_{N-1}]^T,$$

$$r^{(2)} = [0 r'_0, \dots, (N-1) r'_{N-1}]^T,$$

$$R = [r^{(0)}, r^{(1)}, r^{(2)}], \text{ y}$$

$$p = [c', c' \Delta b, c' \Delta a]^T,$$

entonces dada la ecuación (4.12), la actividad relativa y las derivas vienen dadas por $y = R^* g$ según se discutió en el anterior capítulo. Una vez obtenido y , la deriva en umbral se estima como y_1/y_0 y la deriva en ganancia como y_2/y_0 . Cuanto menores sean las desviaciones Δa y Δb en g más se parecerán los vectores y y p , es decir, mejor será la estimación de parámetros.

La discusión anterior puede ser presentada de forma alternativa utilizando el

concepto del jacobiano. Si $\nabla^g(c', a, b)$ es el jacobiano de $g(c', a, b)$, sus tres columnas son

$$\nabla_0^g(c', 1, 0) = r(1, 0),$$

$$\nabla_1^g(c', 1, 0) = c' r'(1, 0),$$

$$\nabla_2^g(c', 1, 0) = c' j r'_j(1, 0), \quad 0 \leq j \leq N-1,$$

por lo que en el caso de pequeña deriva tenemos que

$$\begin{aligned} g(c', 1+\Delta a, 0+\Delta b) &\approx g(c', 1, 0) + \nabla^g(c', 1, 0) [0, \Delta a, \Delta b]^T \\ &= c' r + c' \left[\frac{\nabla_1^g(c', 1, 0)}{c'}, \frac{\nabla_2^g(c', 1, 0)}{c'} \right] [\Delta a, \Delta b]^T \\ &= \left[r, \frac{\nabla_1^g(c', 1, 0)}{c'}, \frac{\nabla_2^g(c', 1, 0)}{c'} \right] [c', c' \Delta a, c' \Delta b]^T \end{aligned}$$

Definiendo

$$M \equiv \left[r, \frac{1}{c'} \nabla_1^g(c', 1, 0), \frac{1}{c'} \nabla_2^g(c', 1, 0) \right]^*$$

tenemos que $p \approx M g$. De esta forma se elimina de M la dependencia de c' y por lo tanto la dependencia de M respecto al espectro problema particular g . M se construye a partir de la referencia r mediante un algoritmo adaptativo y almacena la información adecuada para estimar, además de la contribución relativa, un par adicional de parámetros mediante una asociación simple. El vector respuesta se recupera mediante tres productos internos y dos divisiones. Estas características permiten considerar el operador ampliado M como una *memoria asociativa*.

Es necesario, por otra parte, conocer el rango de deriva, tanto de umbral como de

ganancia, que hace válida la expansión lineal de la mezcla según la ecuación (4.12), es decir, al hablar de deriva, interesa conocer qué es lo que se entiende por "pequeña". Para ello, hemos sometido un espectro simulado a distorsiones diversas en umbral y ganancia y hemos recogido el valor absoluto del error cometido en la estimación de las mismas, así como el error global de la regresión. Considerando que $y = Mg$, definimos el error en la estimación de la deriva en umbral como

$$\varepsilon_U \equiv \left| \Delta b - \frac{y_1}{y_0} \right|$$

y el error en la estimación en ganancia como

$$\varepsilon_G \equiv \left| \Delta a - \frac{y_2}{y_0} \right|.$$

Las Fig. 4.6, Fig. 4.7 y Fig. 4.8 muestran los valores obtenidos de ε_U , ε_G y $sen(\alpha)$ respectivamente al aplicar las derivas indicadas al espectro de la Fig. 4.3.

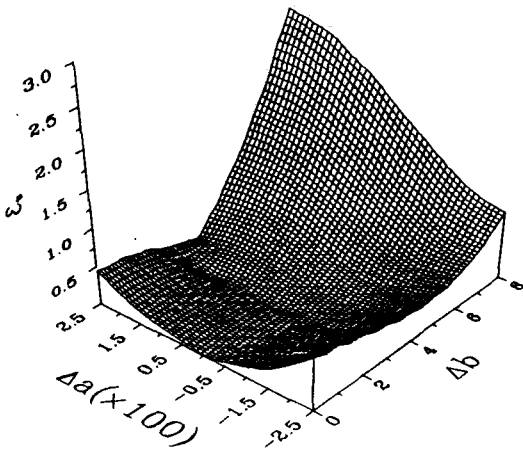


Fig. 4.6 Error en la estimación de la DERIVA en UMBRAL en función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3.

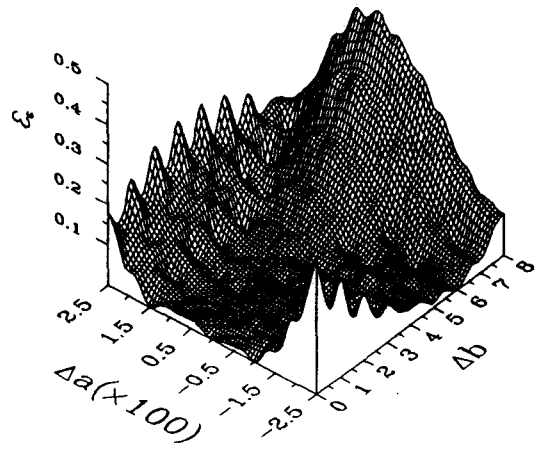


Fig. 4.7 Error en la estimación de la DERIVA en GANANCIA en función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3.

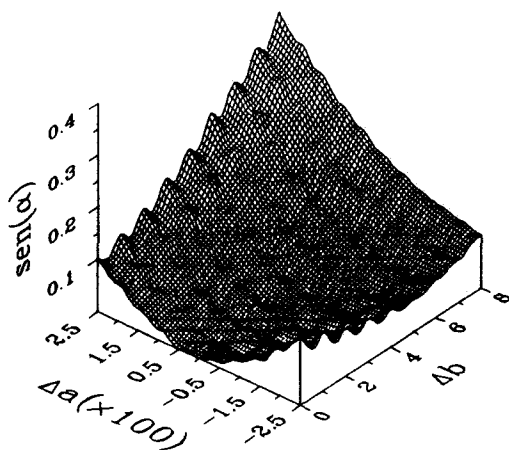


Fig. 4.8 BONDAD del ENCAJE como función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3.

4.2.2 DERIVAS EN MEZCLAS

Según la primera expresión de (4.11), el espectro de emisión de un sólo radionúclido tomado en condiciones de deriva puede ser descrito por el espectro de referencia correspondiente, escalado según la actividad de la muestra y, por último, modificado de acuerdo al par de parámetros Δa y Δb .

Observemos ahora la Fig. 4.9. Ilustra el caso en el cual al espectro problema contribuyen dos fuentes de referencia diferentes $r_1(x)$ y $r_2(x)$, y, además, la captura de la mezcla se ha llevado a cabo en condiciones de deriva. La descripción de la misma vendrá dada por:

$$g(x) = c_1(1+\Delta a) r_1[(1+\Delta a)x + \Delta b] + c_2(1+\Delta a) r_2[(1+\Delta a)x + \Delta b] + e(x) \quad (4.13)$$

en virtud del principio de superposición. Puede observarse en esta ecuación que las dos contribuciones relativas son diferentes porque las actividades de las referencias son, en

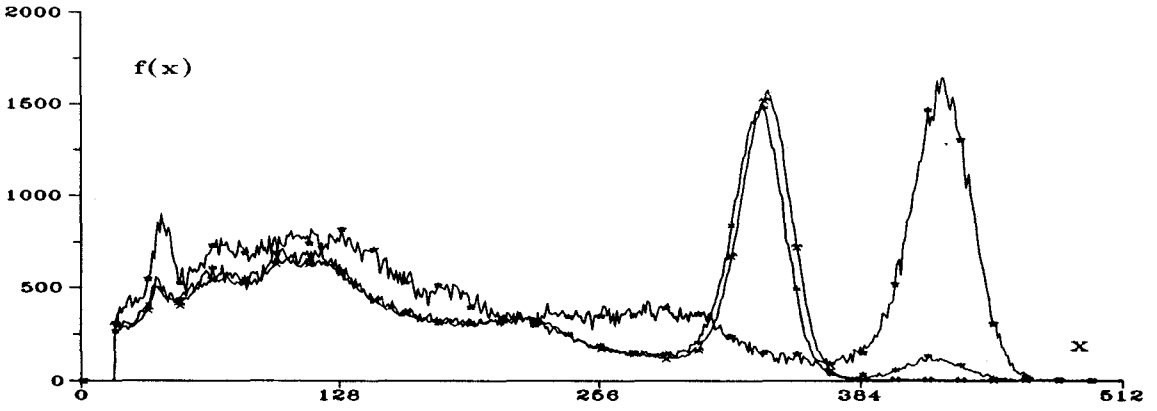


Fig. 4.9 Espectros de dos referencias, ^{137}Cs y ^{54}Mn , y una mezcla de ambos en la que se observa una pequeña deriva instrumental en ganancia.

general, diferentes. Sin embargo, la deriva instrumental es un fenómeno originado en el sistema de tratamiento del impulso eléctrico que genera cada fotón incidente y, por lo tanto, es independiente de la composición de la fuente de radiación. Por este motivo, todas las referencias que participan en la mezcla estarán afectadas por la misma distorsión.

Como en el caso del espectro simple, nos serviremos del concepto de la matriz jacobiana para discutir la descomposición asociativa de un espectro compuesto. Supongamos que se consideran mezclas en las que interviene un máximo de K referencias posibles. De este modo se plantea un problema de estimación de $K+2$ parámetros en la mezcla g . Si $\nabla^g(c', a, b)$ es el jacobiano de $g(c', a, b)$, sus $K+2$ columnas vienen dadas como

$$\nabla_i^g(c', 1, 0) = r_i, \quad 1 \leq i \leq K,$$

$$\nabla_{K+1+j}^g(c', 1, 0) = \sum_{k=1}^K c_k' j r'_{kj}, \quad 0 \leq j \leq N-1,$$

$$\nabla_{K+2}^g(c', 1, 0) = \sum_{k=1}^K c_k' r'_k.$$

por lo que, en el caso de pequeña deriva

$$g(c', 1+\Delta a, 0+\Delta b) \approx g(c', 1, 0) + \nabla^g(c', 1, 0) [0, \Delta a, \Delta b]^T.$$

En el caso multicomponente desgraciadamente no resulta posible extraer el vector c' de la matriz jacobiana $\nabla^g(c', 1, 0)$, lo que se traduce en que tampoco sea posible derivar un operador asociativo de la misma como se hizo en el caso de un único componente. En consecuencia, el problema debe ser reformulado. Vamos a hacerlo introduciendo en el modelo parámetros adicionales. La idea consiste en considerar que cada referencia r_i está sometida a una deriva particular $(\Delta a_i, \Delta b_i)$, con lo cual

$$g(c, a, b) = \sum_{i=1}^K c'_i r_i(a_i, b_i).$$

Si $\nabla^g(c, a, b)$ es el jacobiano de $g(c, a, b)$, en las condiciones de referencia sus $3K$ columnas son

$$\nabla_{3i-3}^g(c', 1, 0) = r_i,$$

$$\nabla_{3i-2-j}^g(c', 1, 0) = c'_i j r'_{i,j}, \quad 0 \leq j \leq N-1,$$

$$\nabla_{3i-1}^g(c', 1, 0) = c'_i r'_i.$$

donde $1 \leq i \leq K$, por lo que en el caso de pequeña deriva

$$g(c', 1+\Delta a, 0+\Delta b) \approx g(c', 1, 0) + \nabla^g(c', 1, 0) [0, \Delta a_1, \Delta b_1, \dots, 0, \Delta a_K, \Delta b_K]^T$$

$$= \sum_{i=1}^K c'_i r_i + [c'_1 \frac{\nabla_1^g}{c'_1}, c'_1 \frac{\nabla_2^g}{c'_1}, \dots, c'_K \frac{\nabla_{3K-2}^g}{c'_K}, c'_K \frac{\nabla_{3K-1}^g}{c'_K}] [\Delta a_1, \Delta b_1, \dots, \Delta a_K, \Delta b_K]^T$$

$$= [r_1, \frac{\nabla_1^g}{c'_1}, \frac{\nabla_2^g}{c'_1}, \dots, r_K, \frac{\nabla_{3K-2}^g}{c'_K}, \frac{\nabla_{3K-1}^g}{c'_K}] [c'_1, c'_1 \Delta a_1, c'_1 \Delta b_1, \dots, c'_K, c'_K \Delta a_K, c'_K \Delta b_K]^T$$

Si se define

$$M \equiv [r_1, \frac{1}{c'_1} \nabla_1^g(c', 1, 0), \frac{1}{c'_1} \nabla_2^g(c', 1, 0), \dots, r_K, \frac{1}{c'_K} \nabla_{3K-2}^g(c', 1, 0), \frac{1}{c'_K} \nabla_{3K-1}^g(c', 1, 0)]^T,$$

como en el caso de espectros simples se elimina de $\nabla^g(c', 1, 0)$ la dependencia de c' y M se constituye como una memoria asociativa independiente de la mezcla problema g . Si se denota $p = [c_1, c_1 \Delta a_1, c_1 \Delta b_1, \dots, c_K, c_K \Delta a_K, c_K \Delta b_K]^T$ entonces $p \approx M g(c', 1 + \Delta a, \Delta b)$.

El valor de Δb puede deducirse tanto de cualquiera de los p_{3i-1} , $1 \leq i \leq K$, y el valor de Δa de cualquiera de los p_{3i-2} , $1 \leq i \leq K$. Si bien estas estimaciones siempre debieran ser iguales, ya que es una restricción física de la formación del espectro, análisis prácticos de espectros reales han mostrado que la precisión de estos valores está fuertemente ligada a la estadística del espectro. El criterio de elección adecuado es estimar la deriva a partir del elemento de referencia i con mayor valor de c'_i , ya que este es el elemento menos afectado por el ruido gaussiano que se superpone a la mezcla.

La ventaja fundamental de esta forma de estimar pequeñas derivas radica en que permite introducir un resolutor lineal, la matriz M , que puede ser implementada en tecnología VLSI u optoelectrónica con el propósito de analizar espectros en tiempo real sujetos a derivas instrumentales ([2], [9]). El precio pagado es la cantidad de parámetros añadidos que han de ser estimados, ya que prácticamente multiplica por tres el número real de los mismos. Así, en un modelo de K referencias $K+2$ parámetros físicos deben ser estimados a partir de $3K$ parámetros formales.

La ampliación de la memoria mejora la calidad del análisis de la mezcla de forma considerable. En el capítulo anterior se estableció como criterio de valoración de la bondad de un análisis la magnitud del ángulo que forma la mezcla problema con el subespacio expandido por el conjunto de referencias. Como veremos en el siguiente apartado, el seno de este ángulo disminuye sensiblemente con la ampliación de las referencias mejorando la bondad del análisis. Resulta ilustrativo a este respecto el apreciar cómo difieren los residuos

$e = My - x$ producto del análisis de una mezcla x cuando se contemplan derivas y cuando no se hace. La Fig. 4.9 mostraba una mezcla de ^{137}Cs y ^{54}Mn afectada por una deriva pequeña. La Fig. 4.10 muestra el residuo generado cuando se presenta a la memoria no ampliada M construida con las referencias de ^{137}Cs y ^{54}Mn representadas junto a ella en Fig. 4.9.

Las Fig. 4.11 y Fig. 4.12 contienen los residuos respectivos de esta misma mezcla tras ampliar la memoria al primer y segundo orden de Taylor respectivamente. El desplazamiento del pico principal de la mezcla respecto de la referencia que contiene este pico se detecta inmediatamente al observar la Fig. 4.10. En ella, la estructura de corte abrupto del

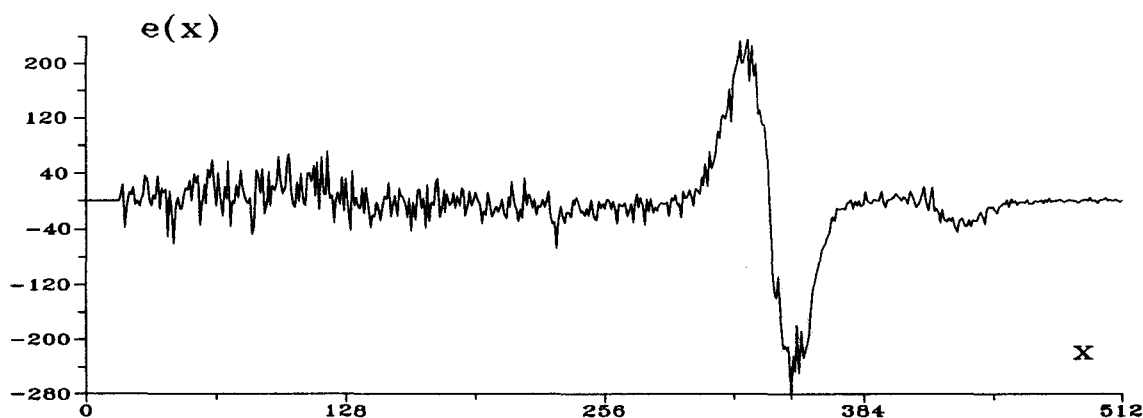


Fig. 4.10 Error de predicción proporcionado por una memoria no ampliada.

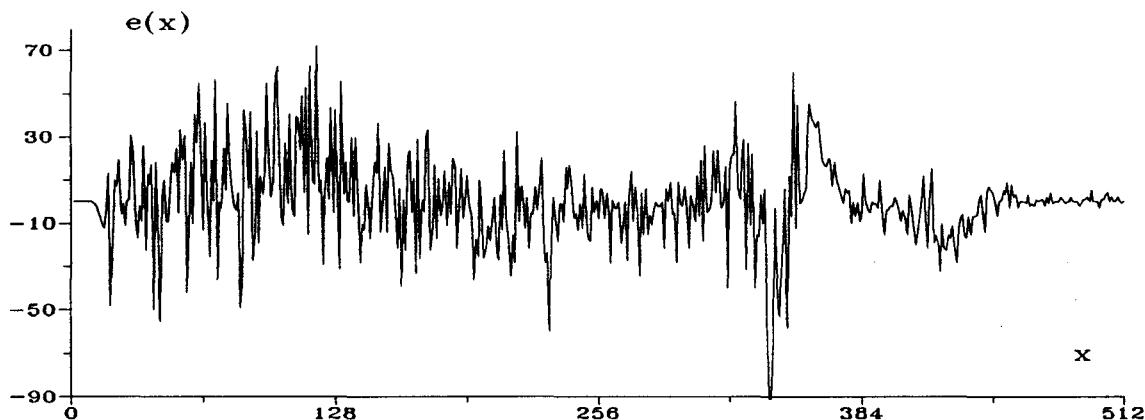


Fig. 4.11 Error de predicción proporcionado por una memoria ampliada al primer orden de Taylor.

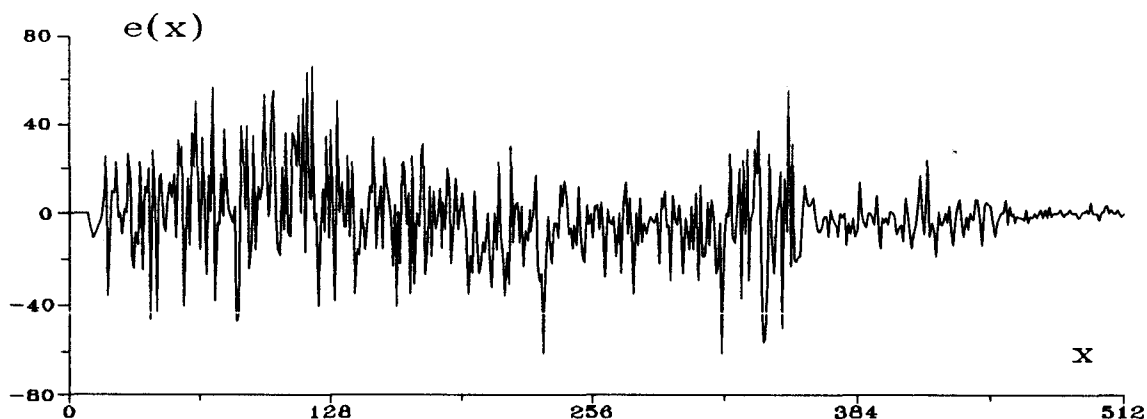


Fig. 4.12 Error de predicción proporcionado por una memoria ampliada al segundo orden de Taylor.

canal 334 se corresponde con la diferencia de dos gaussianas que se solapan y es signo inequívoco de derivas. En contraste, las nuevas referencias del modelo expandido absorben estas discrepancias y dan como resultado un residuo prácticamente errático que constituye una validación del modelo, tanto en lo que se refiere a la suficiencia del conjunto de referencia como al tratamiento de la deriva.

4.2.3 DISCUSION EN TORNO A LA INTERPRETACION DE DERIVAS

En el anterior epígrafe se mostró cómo una memoria asociativa ampliada es capaz de descomponer adecuadamente una mezcla afectada por derivas y, por tanto, estimar de forma correcta las contribuciones relativas de sus componentes. El paso siguiente es proporcionar la deriva en ganancia y umbral en términos y unidades precisas, es decir, dar una estimación de la desviación en umbral en *canales* y de la desviación en ganancia en *tanto por ciento*. Diremos que la mezcla $f(x)$ ha sufrido una deriva de u canales cuando se transforma en $f(x - u)$ y una deriva del $g\%$ cuando se transforma en $f(x(1 + (g/100)))$.

4.2.3.1 DERIVA EN UMBRAL

Supongamos un conjunto R de una única referencia que se amplía para estimar deriva en umbral, de modo que $R = [r^{(0)} \mid r^{(1)}]$. Supongamos un espectro problema s_i , $1 \leq i \leq N$, igual al espectro de referencia pero afectado por una desviación en umbral de un canal hacia la derecha (energías mayores). Supongamos asimismo que disponemos de un analizador multicanal con un número de canales N suficientemente grande como para considerar pequeña la deriva de un sólo canal. La forma analítica del espectro problema es entonces $s_i = r_{i-1}$, $1 \leq i \leq N$, $s_0 = 0$. Al presentarlo a la memoria por medio de la operación algebraica Rs , obtenemos el vector $p^T = [c, \Delta\beta]$. Estimemos $r^{(1)}$ como la primera diferencia de $r^{(0)}$, de modo que

$$r_i^{(1)} = r_i^{(0)} - r_{i-1}^{(0)}. \quad (4.14)$$

Teniendo en cuenta que $r_{i-1}^{(0)} = r_i^{(0)} - (r_i^{(0)} - r_{i-1}^{(0)})$, tenemos que $r_{i-1}^{(0)} = r_i^{(0)} - r_i^{(1)}$. Ahora bien, dado que $s_i = r_{i-1}^{(0)}$, finalmente $s_i = r_i^{(0)} - r_i^{(1)}$. Según esta ecuación, un espectro con una deriva de un canal es expresable de forma exacta como una expansión lineal del espectro original y su primera diferencia. Esto significa que $R^*s = [1, -1]^T$, es decir, la memoria estima la deriva de forma exacta cambiada de signo. Este problema del signo se soluciona introduciendo en el conjunto de referencia $r^{(1)} = -r'$, en lugar de lo que hasta ahora se había establecido, $r^{(1)} = +r'$. Por supuesto, aunque la deriva no sea exactamente de un canal, su estimación sigue siendo correcta bajo la suposición de que la misma es pequeña.

4.2.3.2 DERIVA EN GANANCIA

Supongamos un espectro problema s_i , $1 \leq i \leq N$, igual al espectro de referencia pero afectado por una desviación en ganancia del 2%. La forma analítica del espectro problema es entonces $s(x) = r((1+0.02)x)$. Al presentarlo a la memoria por medio de la

operación algebraica Ms , obtenemos el vector $p^T = [c, \Delta a]$. donde Δa es un valor en torno al 0.02. Conviene que la memoria exprese la ganancia en tanto por ciento de modo que el valor obtenido no sea 0.02 sino 2. Esto conlleva que las estimaciones de umbral y ganancia estén dentro del mismo orden de magnitud. Esta consideración no es caprichosa, sino que responde a la necesidad de que las normas de todas las filas de la memoria no difieran en exceso. Dado un espectro de referencia r , la referencia ampliada que estima la deriva en umbral es, como sabemos, r'_i , y la que estima la deriva en ganancia ir'_i , $0 \leq i \leq N-1$ o i' . Como es fácil comprender, los valores de esta segunda referencia ampliada crecen rápidamente al aumentar i , provocando una alta discrepancia entre las normas de r'_i , y i' . El efecto sobre la norma de un vector en la operación de pseudoinversión es equivalente al efecto del valor absoluto de un número en la división de números reales. Números grandes tienen inversos pequeños y viceversa. Como consecuencia, las filas de la memoria correspondientes a los vectores i' , tienen una norma muy pequeña que originan resultados pequeños respecto a la unidad como 0.02. Si dividimos por 100 al vector i' , es fácil demostrar que su fila correspondiente en la pseudoinversa queda multiplicada por este mismo valor, dejando las dos filas restantes inalteradas. Con esta maniobra se alcanzan dos objetivos a la vez. Por una parte, la memoria interpreta la deriva en ganancia directamente en tanto por ciento y, por otra, reducimos el rango dinámico de las referencias, lo cual redundará, lógicamente, en una mayor precisión del algoritmo de pseudoinversión.

4.3 EL ORDEN DE LA EXPANSION DE TAYLOR

La ecuación (4.8) puede ampliarse añadiendo órdenes superiores al desarrollo. Cuanto mayor sea el orden, mayores desviaciones de ganancia y umbral seremos capaces de analizar de modo fiable. Ahora bien, los algoritmos de cómputo de pseudoinversa presentados en el capítulo anterior confían en la independencia lineal del conjunto ampliado de referencias. Considerando espectros N -dimensionales, el máximo de vectores en R está teóricamente limitado a N referencias. Esto significa que al desdoblar cada espectro de

referencia en tres para estimar derivas, el límite de N referencias puras queda reducido a $N/3$. Si desdoblamos en cinco, a $N/5$, etc. Por otra parte, el aumento del número de columnas en R provoca una degeneración del índice de condicionamiento de $R^T R$, lo cual provoca un aumento en el error del vector $R^+ g$. Aparece así un compromiso en el plano teórico que impondrá restricciones en la práctica. Hechas estas consideraciones, vamos a ver que la forma analítica de los desarrollos sucesivos de un espectro responden a una regla recursiva sencilla.

4.3.1 DESARROLLOS SUPERIORES AL PRIMERO

El desarrollo de $f(ax + b)$ en $(a = 1, b = 0)$ podemos describirlo como:

$$f(ax + b) = f(x) + O_x^{(1)}(a, b) + O_x^{(2)}(a, b) + O_x^{(3)}(a, b) + \dots$$

donde

$$O_x^{(1)}(a, b) = x \frac{df}{dx}(a-1) + \frac{df}{dx}(b-1)$$

$$= \nabla_x f[\Delta a, \Delta b]^T$$

$$O_x^{(2)}(a, b) = \frac{1}{2!} \left[\frac{\delta^2 f}{\delta a^2} (a-1)^2 + 2 \frac{\delta^2 f}{\delta a \delta b} (a-1)(b-0) + \frac{\delta^2 f}{\delta b^2} (b-0)^2 \right]$$

$$= \frac{1}{2} [\Delta a, \Delta b] \nabla_x^2 f[\Delta a, \Delta b]^T,$$

(donde a su vez $\nabla_x^2 f$ es la matriz hessiana de la función $f(x, a, b)$ respecto de a y b),

$$O_x^{(3)}(a,b) = \frac{1}{3!} \left[\frac{\delta^3 f}{\delta a^3} (a-1)^3 + 3 \frac{\delta^3 f}{\delta a^2 \delta b} (a-1)^2 (b-0) + \right. \\ \left. + \frac{\delta^3 f}{\delta b^3} (b-0)^3 + 3 \frac{\delta^3 f}{\delta a \delta b^2} (a-1)(b-0)^2 \right]$$

Los términos de $O_x^{(1)}(a, b)$, $f'(x)$ y $xf'(x)$ fueron obtenidos en el anterior apartado. Los términos correspondientes a órdenes superiores exigen un prolijo desarrollo que evitamos al lector. Para $O^{(2)}(a, b)$ resultan ser:

$$\frac{\delta^2 f}{\delta a^2} = x^2 f''(x),$$

$$\frac{\delta^2 f}{\delta a \delta b} = x f''(x),$$

$$\frac{\delta^2 f}{\delta b^2} = f''(x),$$

y para $O^{(3)}(a, b)$:

$$\frac{\delta^3 f}{\delta a^3} = x^3 f'''(x),$$

$$\frac{\delta^3 f}{\delta a^2 \delta b} = x^2 f'''(x),$$

$$\frac{\delta^3 f}{\delta a \delta b^2} = x f'''(x),$$

$$\frac{\delta^3 f}{\delta b^3} = f'''(x).$$

De los resultados anteriores puede inferirse como regla general que para expresar $O^{(i)}(a, b)$ necesitamos las derivadas $f^{(i)}(x)$, $xf^{(i)}(x)$, $x^2f^{(i)}(x)$, ..., $x^{i-1}f^{(i)}(x)$ y $x^if^{(i)}(x)$. La expansión de un espectro de referencia r_j hasta un determinado orden de expansión e puede sistematizarse si las expresiones determinadas anteriormente $x^mr^{(n)}(x)$, donde $1 \leq m \leq n$ y $1 \leq n \leq e$, se ordenan de acuerdo a algún criterio. Uno posible es el que sigue:

1. $r_j^{(0)} \Leftrightarrow r_j(x)$,
2. $r_j^{(1)} \Leftrightarrow r'_j(x)$,
3. $r_j^{(2)} \Leftrightarrow x r'_j(x)$,
4. $r_j^{(3)} \Leftrightarrow r''_j(x)$,
5. $r_j^{(4)} \Leftrightarrow x r''_j(x)$,
6. $r_j^{(5)} \Leftrightarrow x^2 r''_j(x)$,
7. $r_j^{(6)} \Leftrightarrow r'''_j(x)$,
8. $r_j^{(7)} \Leftrightarrow x r'''_j(x)$,
9. $r_j^{(8)} \Leftrightarrow x^2 r'''_j(x)$,
10. $r_j^{(9)} \Leftrightarrow x^3 r'''_j(x)$,
11. $r_j^{(10)} \Leftrightarrow r_j^{(IV)}(x)$,
12. $r_j^{(10)} \Leftrightarrow x r_j^{(IV)}(x)$, etc.

Es intuitivo pensar que un aumento del orden de Taylor utilizado en la construcción de la memoria posibilitaría una estimación de derivas mayores. Las figuras siguientes (Fig. 4.13 a Fig. 4.18) muestran los errores de la estimación los valores de umbral y de ganancia introducidos en el espectro de la Fig. 4.3, así como el factor de ajuste cuando se amplían a seis y a nueve las referencias consideradas por elemento respectivamente. Dejamos que el lector saque sus propias conclusiones, si bien puede apreciarse que la mejora es notable.

Es importante no perder de vista que las sucesivas derivadas de las referencias deben obtenerse mediante métodos numéricos a partir de datos discretos. La teoría sobre este particular puede encontrarse en los trabajos [3], [4], [5], [7], [10], y [13]. Hay que tener en cuenta que una derivación no es más que un tipo de filtrado paso alto, por lo tanto un extractor de altas frecuencias y, como tal, magnificador de ruido. Este ruido se propaga y amplifica en derivaciones sucesivas acrecentando la corrupción de las mismas. A pesar de todo, la mejora progresiva obtenida en la calidad de las estimaciones al aumentar el orden de Taylor queda evidenciada en las figuras que siguen, que comparan los errores de estimación encontrados sobre un orden de expansión de seis y otro orden de expansión de nueve.

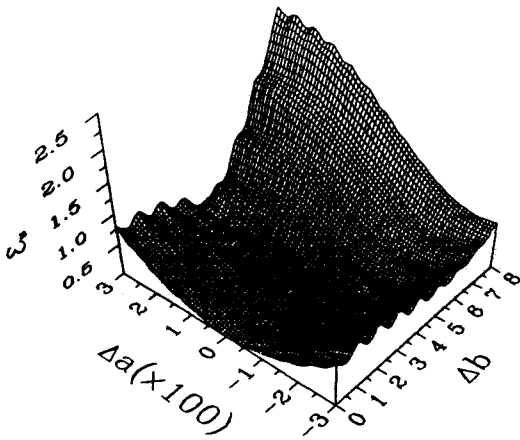


Fig. 4.13 Error en la estimación de la DERIVA en UMBRAL presente en el espectro de la Fig. 4.3. Seis términos de Taylor fueron considerados.

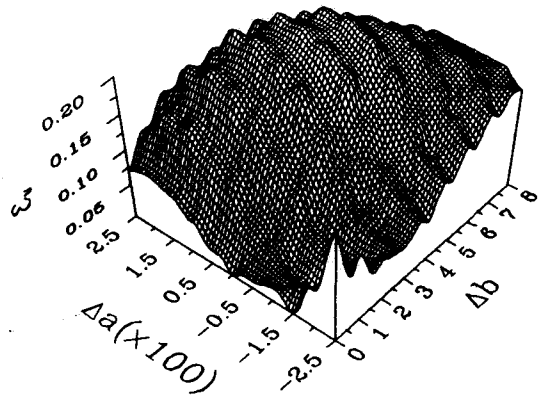


Fig. 4.14 Error en la estimación de la DERIVA en UMBRAL presente en el espectro de la Fig. 4.3. Nueve términos de Taylor fueron considerados.

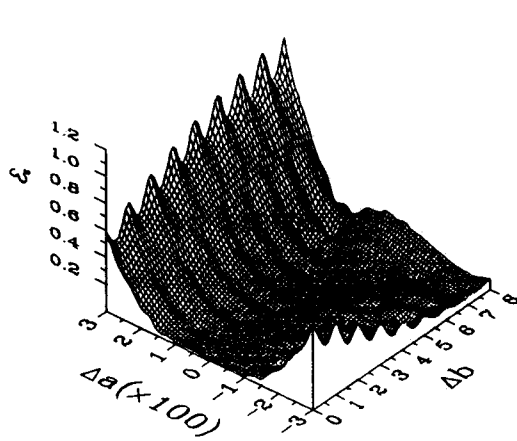


Fig. 4.15 Error en la estimación de la DERIVA en GANANCIA en el espectro de la Fig. 4.3. Seis términos de Taylor fueron considerados.

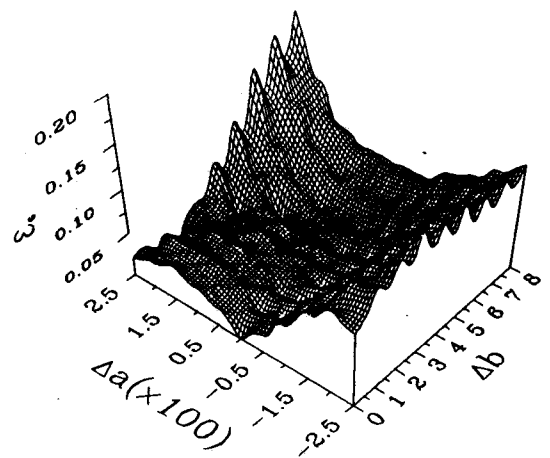


Fig. 4.16 Error en la estimación de la DERIVA en GANANCIA del espectro de la Fig. 4.3. Nueve términos de Taylor fueron considerados.

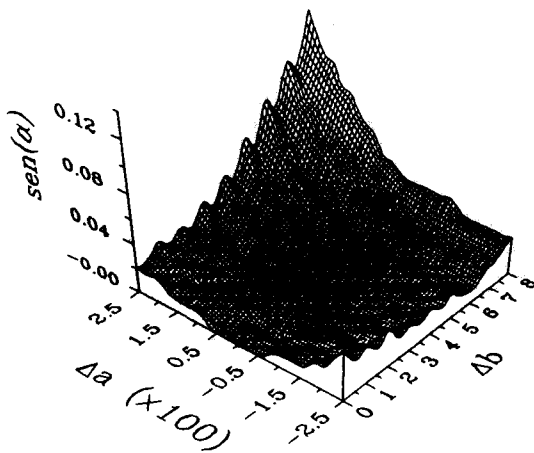


Fig. 4.17 BONDAD del ENCAJE en función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3. Seis términos de Taylor fueron considerados.

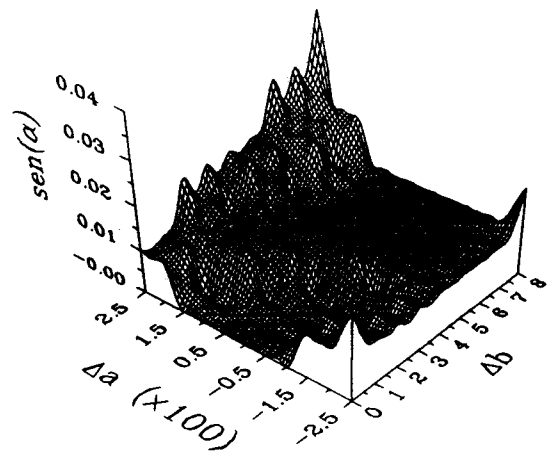


Fig. 4.18 BONDAD del ENCAJE en función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3. Nueve términos de Taylor fueron considerados.

4.3.2 REDUCCION DE DATOS

Las tres gráficas siguientes Fig. 4.19, Fig. 4.20 y Fig. 4.21 muestran los errores obtenidos en el análisis de una mezcla sintética obtenida al combinar cinco referencias de un conjunto de referencia de siete $[r_1, r_2, r_3, r_4, r_5, r_6, r_7]$. La mezcla problema x responde a la siguiente combinación:

$$x = 0.1 \times r_1 + 1.5 \times r_2 + 1.0 \times r_3 + 0.5 \times r_4 + 1.0 \times r_5 \quad (4.15)$$

Las derivas en umbral y ganancia fueron estimadas a partir de la segunda referencia por ser la que proporcionaba una mayor contribución como ya se discutió anteriormente. Tanto la mezcla como las referencias son vectores de 128 componentes en lugar de los 1024 tratados hasta ahora. Esta reducción de datos de ocho a uno ha sido realizada tomando el valor medio de las cuentas acumuladas en ocho canales contiguos. En esta experiencia citada y en todas las que se han llevado a la práctica los índices de error no varían de forma sustancial respecto a los obtenidos a partir de espectros no compactados de 1024 canales, tal y como demuestran las siguientes figuras. Este hecho trae como consecuencia práctica que el tamaño de la memoria se reduzca en factor de ocho, lo cual facilita la implementación VLSI de la misma así como la de los algoritmos de adaptación discutidos en el tercer capítulo.

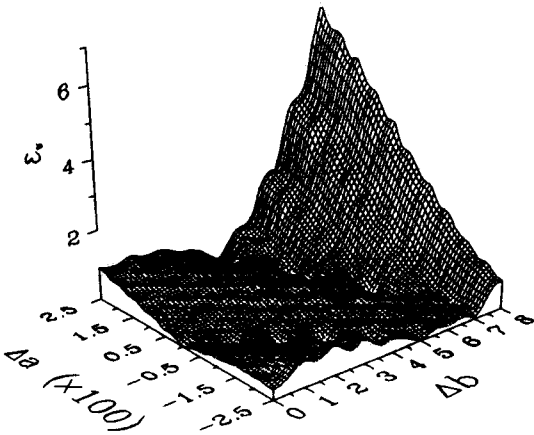


Fig. 4.19 Estimación de la DERIVA en UMBRAL presente tras introducir derivas en la mezcla (4.15). Seis términos de Taylor fueron considerados.

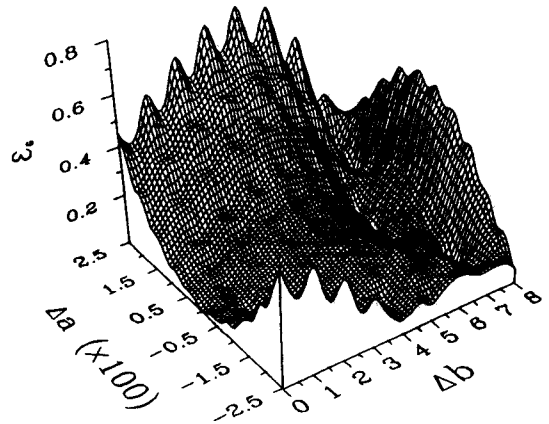


Fig. 4.20 Estimación de la DERIVA en GANANCIA presente tras introducir derivas en la mezcla (4.15). Seis términos de Taylor fueron considerados.

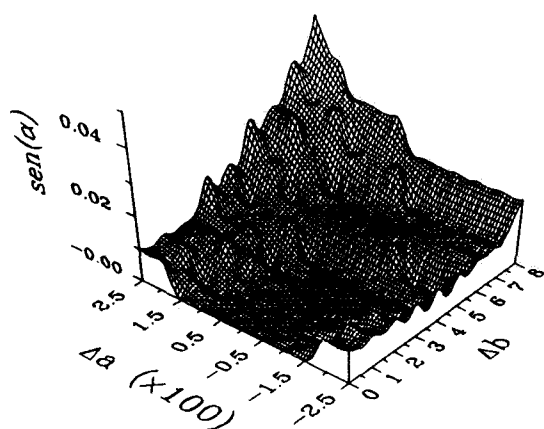


Fig. 4.21 BONDAD del ENCAJE de la mezcla (4.15) tras introducir en la misma derivas diversas. Seis términos de Taylor fueron considerados.

4.3.3 CALCULO NUMERICO DE LOS DESARROLLOS

El tratamiento de las derivas instrumentales impone una ampliación del conjunto de referencia con las derivadas de los espectros base. De este modo, el problema que se plantea es estimar un muestreo de N puntos o vector N -dimensional r' de la derivada $r'(x)$ de una función $r(x)$ que, a su vez, es conocida como un muestreo de N puntos o vector N -dimensional r . Más concretamente, es preciso determinar un operador D de diferenciación de datos discretos tal que $r' = D[r]$.

El método de diferenciación numérica utilizado ha sido el de convolución con un filtro digital cuyos coeficientes han sido tomados de las tablas de Savitzky-Golay ([10], corregidas posteriormente en [13]). Este método es una generalización del concepto de filtrado de media de un conjunto de datos numéricos y es directamente aplicable a la diferenciación de los mismos. La idea básica consiste en tomar una ventana en los datos de $2m + 1$ puntos, ajustar a los datos seleccionados un polinomio de un orden determinado por el método de los mínimos cuadrados y asignar al punto central de la ventana el valor del polinomio en este punto. Como derivada se toma el valor de la derivada del polinomio de ajuste. La aportación de Savitzky y Golay fue la demostración de que el valor central era

evaluable a través de una ponderación de los datos con un vector de pesos independiente de la posición de la ventana. Ellos sistematizaron esta idea proponiendo un conjunto de filtros digitales cuyos coeficientes d_i tabularon para los diversos parámetros implicados: grado del polinomio de ajuste, número m de datos de la ventana y orden de la derivación. La descripción matemática del proceso es la siguiente convolución:

$$r_j = \sum_{i=-m}^{i=m} d_i r_{j+i}$$

Nosotros hemos utilizado los coeficientes d_i de la tabla III de [10] para nueve puntos de datos con el fin de estimar la primera derivada. Para las derivadas sucesivas hemos aplicado el mismo filtro.

4.3.4 EL ESPECTRO DE FONDO EN DETECTORES GAMMA

En ausencia de emisión radiactiva, el analizador multicanal acumula un espectro producto de las cargas que saltan a la banda de conducción no por efecto de una alteración energética provocada por un fotón incidente, sino por efecto estadístico de la temperatura. Los canales correspondientes a energías bajas son, lógicamente, los que más cuentas acumulan. Debido a que este fenómeno se atenúa exponencialmente con la energía, el espectro de fondo tiene así una forma típica:

$$f(x) = K e^{-Cx}$$

La derivada de una exponencial es otra exponencial. En este caso:

$$f^{(i)}(x) = (-1)^i C^i K e^{-Cx}$$

de modo que las diversas derivadas son iguales al espectro de referencia salvo factores

constantes. Si bien el espectro real tomado como función del canal no responde de forma exacta a la exponencial, derivadas muy parecidas llevan la matriz a un estado cercano a la singularidad que puede causar grandes errores numéricos en los algoritmos ya presentados de cálculo de la memoria. Hay que decir que la no expansión del espectro de fondo ha representado una solución satisfactoria en todos los casos prácticos acometidos.

4.4 BUSCANDO EL OPERADOR MINIMO

En el capítulo anterior se planteó la necesidad de que la formación adaptativa de la memoria asociativa se llevase a cabo de un modo dependiente de la mezcla problema. La incorporación exhaustiva de todas las referencias elementales disponibles al algoritmo de pseudoinversión conduce a un resolutor de tamaño excesivo, inexacto y de elaboración lenta. En contraste, la introducción de las referencias apropiadas para descomponer la mezcla y sólo éstas conduce a un operador mínimo más preciso, ágil y de formación más rápida. Huelga decir que la búsqueda del operador mínimo se hace más apremiante si cabe con el aumento de dimensionalidad del problema que supone el tratamiento de las derivas instrumentales. La cuestión que se plantea tras la introducción de las derivas en el modelo de espectro es si el algoritmo adaptativo dependiente de la mezcla problema propuesto en el apartado 3.7 continúa comportándose de forma correcta ante una mezcla problema sin deriva excesiva sin necesidad de introducir en el mismo elementos incompatibles con el modelo original (3.28). Experiencias realizadas sobre un entorno i386-MsDOS con tanto espectros sintéticos como espectros reales de INa(Tl) tomados con un analizador multicanal MCA Camberra 35 Plus en laboratorios del C.I.E.M.A.T. permiten afirmar que el algoritmo de regresión por pasos (3.28) propuesto en el anterior capítulo para mezclas sin deriva continúa vigente en su forma básica aplicado al caso de mezclas afectadas por pequeñas derivas. La versión del mismo que aborda este problema es la que a continuación se propone:

```

 $K = \{\phi\};$ 
 $e := x;$ 
DO
{
 $\rho_i = \cos^2(r_i, e);$ 
Sea  $g / \rho_g = \max_i \rho_i;$ 
 $K := K \cup \{g\};$ 
DO
{
 $R := [\dots | r_i^{(0)} | r_i^{(1)} | r_i^{(2)} | \dots], i \in K;$ 
 $M := R^*;$ 
 $c := Mx;$ 
 $N := \{i / c_i^{(0)} \leq NivelSig\};$ 
 $K := K - \{N\};$ 
} WHILE ( $N \neq \{\phi\}$ );
 $e := [I - RM]x;$ 
 $sen^2(\alpha) := e^T e / x^T x;$ 
} WHILE ( $sen^2(\alpha) < UMBRAL$ );
return ( $M$ );

```

(4.16)

El algoritmo es prácticamente idéntico al original, salvo las sentencias primera y cuarta del bucle "do-while" interno. En la primera modificación se introducen las derivadas de la referencia elegida además de ésta misma. En la segunda modificación se indica que sólo la referencia original y no las derivadas es la considerada en la prueba de rechazo. La experimentación con este algoritmo ha sido positiva incluso con derivas de hasta 8 a 12 canales sobre mezclas de 1024 y ganancias de hasta el 2% sobre el conjunto de referencias sobre el que se trabajó en el capítulo anterior.

4.5 BIBLIOGRAFIA

- [1] Demidovich, B. P., "Problemas y ejercicios de análisis matemático", Ed. Paraninfo, (1982).
- [2] Díaz, J. C. *et al.*, "An Associative Memory to Solve the Mixture Problem in Composite Spectra", 35th Midwest Symposium on Circuits and Systems, Washington, August 9-12, (1992), pp. 891-894.
- [3] Edwards, T. H. *et* Willson, P. D., "Digital Smoothing of Spectra", Applied Spectroscopy, 28, (1974), pp. 541-545.
- [4] Enke, C. G. *et* Nieman, T. A., "Signal-To-Noise Ratio Enhancement By Least Squares Polynomial Smoothing", Analytical Chemistry, Vol. 48, No. 8, (1976), pp. 705-712.
- [5] Gorry, P. A., "General Least-Squares Smoothing and Differentiation By the Convolution (Savitzky-Golay) Method", 62, (1990), pp. 570-573.
- [6] Heath, R. L., "Data Analysis Techniques for Gamma-Ray Scintillation Spectrometry", Atomic Energie Comission, Report IDO-16784, (1962).
- [7] Khan, A., "Problems of Smoothing and Differentiation of Data By Least Squares and Possible Solutions", Analytical Chemistry, 59, (1987), pp. 654-657.
- [8] Parr, R. M. *et* Lucas, H. F., "A Rigorous Least Squares Analysis of Complex Gamma Ray Spectra with Partial Compensation for Instrumental Instability", IEEE Trans. on Nuclear Science, 349, June, (1964), pp. 349-357.
- [9] Rodellar, V. *et al.*, "A VLSI Arithmetic Unit for a Signal Processing Neural Neural Network", 35th Midwest Symposium on Circuits and Systems, Washington, August 9-12, (1992), pp. 1044-1047.
- [10] Savitzky, A. *et* Golay, M. J., "Smoothing and Diferenciation of Data by Simplified Least Squares Procedures", Analytical Chemistry, 36, (1964), pp. 1627-1639.
- [11] Schonfeld, E. *et al.*, "Determination of Nuclide Concentrations in Solutions Containing Low Levels of Radiativity by Least-Squares Resolution of the Gamma-Ray Spectra", Nuclear

Instrumens and Methods, 45, (1966), pp. 1-22.

- [12] Schonfeld, E., "ALPHA - a Computer Program for the Determination of Radioisotopes by Least-Squares Resolution of the Gamma-Ray Spectra", Nuclear Instruments and Methods, 42, (1966), pp. 213-218.

- [13] Steinier, J. *et al.*, "Comments on Smoothing and Differenciation of Data by Simplified Least Square Procedure", Analytical Chemistry, Vol. 44, No. 11, Sept, (1972), pp. 1906-1909.

Capítulo 5

DERIVAS SEVERAS: MEMORIA ASOCIATIVA CON REALIMENTACION

La transformación funcional de un espectro de radiación $f(x)$ al espectro $(1+\Delta a)f((1+\Delta a)x + \Delta b)$ por efecto de derivas instrumentales es de naturaleza no lineal. Por esta misma razón, el problema inverso, a saber, la estimación de Δa y Δb de un espectro con derivas, fuerza al empleo de técnicas no lineales. En el capítulo anterior examinamos el modelo de pequeña deriva y vimos que ésta podía ser determinada a partir de un resolutor lineal o memoria asociativa en una sola operación de matriz-vector. Este nuevo capítulo se ocupa del caso en el que las derivas sean lo suficientemente importantes como para que el modelo de pequeña deriva sea insuficiente

y el problema deba reformularse en el contexto de la teoría de optimización no lineal. Así, dada una mezcla original cualquiera x de parámetros de composición y deriva desconocidos y recogidos en el vector β , la estimación de los mismos conduce al problema de la minimización de $\|x - f(\beta)\|^2$ respecto a β , donde la dependencia de f respecto a β es de carácter no lineal. Esta expresión cuadrática hace que a este tipo de problemas se les conozca como de *regresión no lineal* u *optimización cuadrática no lineal*. A lo largo del presente capítulo, primero, se examinan de un modo general los algoritmos fundamentales de este área de la teoría de optimización, destacando entre los mismos el algoritmo de Gauss-Newton por su estrecha relación con el formalismo de memoria asociativa y, segundo, a partir de una nueva formulación del problema de la descomposición no lineal de un espectro mezcla, se introduce un nuevo algoritmo de ajuste mucho menos costoso computacionalmente que el Gauss-Newton cuyo núcleo es una memoria asociativa ampliada según fue definida en el anterior capítulo. Este procedimiento supone un modelo de operación de la memoria de tipo recursivo de tal forma que la mezcla problema es continuamente presentada a la memoria y corregida en sus parámetros no lineales hasta lograr la minimización de la función de error. Con este modo de empleo, la capacidad del operador M para descomponer mezclas con derivas severas se acrecienta sensiblemente.

5.1 OPTIMIZACION CUADRATICA NO LINEAL

Podemos definir la optimización como la parcela de la matemática que da solución a un problema bien definido que surge a menudo al construir modelos de la realidad física: el de determinar el valor o valores de β que minimizan una función $E(\beta)$ en un dominio dado. Antes de 1940 se conocía poco sobre la optimización de funciones de muchas variables. Se habían llevado a cabo algunos cálculos de mínimos cuadrados y puesto en marcha algunos procedimientos de descenso tipo gradiente en algunos problemas físicos. También se conocía el método de Newton para muchas variables pero, en general, cualquier problema de alguna complejidad -muchas variables- era inabordable. No fue sino hasta la llegada del computador cuando se produjo la revolución de los métodos de optimización y, en general, del análisis numérico. Los años cuarenta y cincuenta vieron el nacimiento y

desarrollo de la programación lineal, pero el hecho que marcó el despegue de los métodos de optimización fue el trabajo de Davidon ([4]) en 1959 sobre la técnica de métrica variable que, frente a los métodos previos, proporcionaba una estabilidad numérica que permitía acometer la optimización de funciones de muchas variables.

Existe toda una variedad de métodos de optimización cuadrática no lineal ([3], [6], [7], [8], [9], [10], [13]) pero, de una forma u otra, todos ellos son variaciones del algoritmo intuitivo de descenso por la función de error en el espacio de los parámetros del modelo a partir de una estimación inicial cercana a la óptima. Este último punto es de una importancia decisiva ya que, a diferencia de los problemas de estimación lineal, la función de error no es en general de forma parabólica y, en consecuencia, puede presentar mínimos locales. Por este motivo, la solución proporcionada por todos los algoritmos de descenso está condicionada por la estimación inicial de parámetros escogida.

De todos los problemas que trata la teoría de optimización vamos a concentrarnos en aquel que resulta del ajuste *no restringido* de un conjunto de observaciones experimentales (un espectro de cuentas) a un modelo paramétrico de formación del mismo. La técnica usualmente utilizada en este contexto es la de minimizar la forma cuadrática no lineal que resulta al evaluar el error del ajuste. A continuación se introduce formalmente el citado problema.

5.1.1 FORMULACION GENERAL DEL PROBLEMA DE REGRESION NO LINEAL

Sean N observaciones x_i para las cuales se proporciona el siguiente modelo de descripción:

$$x_i = f_i(\beta_1, \dots, \beta_p) + e_i, \quad 0 \leq i \leq N-1, \quad (5.1)$$

o bien, en notación vectorial,

$$x = f(\beta) + e \quad (5.2)$$

donde β es un vector de parámetros a determinar, f_i una función no lineal en β y e el ruido asociado a la observación ([13]). El problema que se plantea es el de hallar los parámetros que minimizan la norma del error cuadrático:

$$\underset{(\beta)}{\text{Minimizar}} \quad E(\beta) \equiv \|x - f(\beta)\|^2$$

La función $E(\beta)$ es un paraboloide cuando f_i es lineal en β y los contornos que determina $E(\beta)$ constante son elipsoides. A medida que aumenta la alinealidad en β , las formas elípticas se deforman pero, aún así, estos contornos son prácticamente elípticos en la vecindad del mínimo, lo que significa que la suposición de linealidad local es una aproximación razonable. Esta fue la hipótesis de trabajo del capítulo anterior, en la que los parámetros no lineales presentes en una mezcla, es decir, la deriva en umbral y ganancia, eran pequeños. En el capítulo presente consideramos que la suposición ya no es válida, por lo que se hace necesario recurrir a los métodos que tradicionalmente se han venido empleando en la resolución de problemas no lineales. A fin de alcanzar una perspectiva global de los mismos, se examinan a continuación los más representativos.

5.1.2 REVISION DE METODOS

Consideremos el caso unidimensional, con una observación x y una función $f(\beta)$ (Fig. 5.1a). El problema es encontrar el parámetro β_* que haga mínima la diferencia entre x y $f(\beta)$ que puede plantearse formalmente como

$$\underset{(\beta)}{\text{Minimizar}} \quad E(\beta) \equiv [x - f(\beta)]^2$$

Según se aprecia en la Fig. 5.1c, la función de error cuadrático $[x - f(\beta)]^2$ puede tener mínimos locales. Por esta razón, cualquier método construido sobre la idea del

descenso progresivo en la función de error, como es el caso de los de minimización cuadrática no lineal, lleva a soluciones locales. Los tres métodos fundamentales, de los cuales se derivan todos los demás, son el método del gradiente, el método de Newton, y el método de Gauss-Newton.

5.1.2.1 EL METODO DEL GRADIENTE

El gradiente de la función de error $E(\beta)$ viene dado por la expresión $\nabla^E(\beta) = -2 \nabla^f(\beta) [f(\beta) - x]$, donde $\nabla^f(\beta)$ es el jacobiano de $f(\beta)$. Podemos construir métodos basados en el descenso por la función de error siguiendo la dirección del gradiente negativo. Por ejemplo, el *algoritmo del descenso más pronunciado*

$$\beta_{n+1} = \beta_n + \lambda_n \nabla^f(\beta_n) [x - f(\beta_n)] . \quad (5.3)$$

Como puede apreciarse, la carga computacional de cada iteración es relativamente baja ya que no es necesario llevar a cabo ninguna inversión de matrices, concentrándose ésta en la evaluación de $\nabla^f(\beta)$ y $f(\beta)$ en cada una de ellas. Sin embargo, el número de iteraciones es generalmente alto.

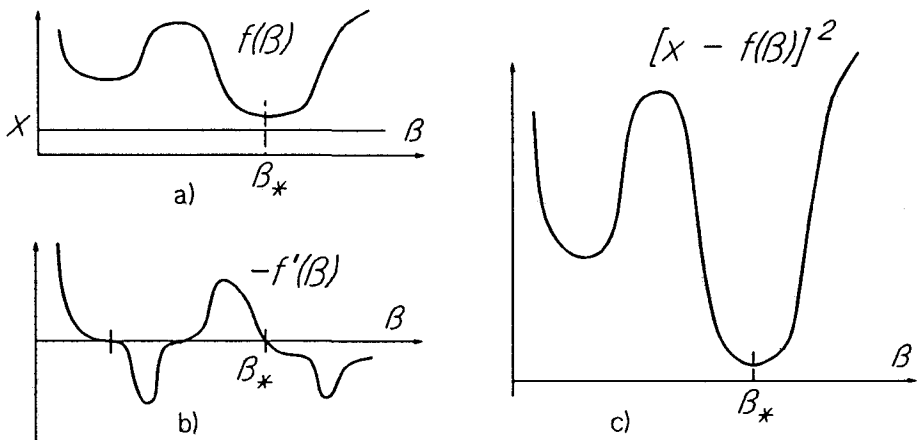


Fig. 5.1 La función objetivo $E(\beta)=[x - f(\beta)]^2$ presenta, en general, mínimos locales cuando $f(\beta)$ es no lineal en β .

5.1.2.2 EL METODO DE GAUSS-NEWTON

Los métodos basados en la expansión en serie de Taylor (a veces llamado el método de Gauss o de Gauss-Newton) emplean la aproximación lineal que proporcionan los primeros términos del desarrollo de Taylor de las funciones $f_i(\beta)$ en las cercanías de β . Sabemos que en una vecindad apropiada puede llevarse a cabo la aproximación

$$f(\beta + \Delta\beta) \approx f(\beta) + [\nabla^f(\beta)] [\Delta\beta]$$

donde ∇^f es la matriz jacobiana del vector de funciones f . Así, si el conjunto x de observaciones viene caracterizado por un vector de parámetros relativamente cercano a β_0 puede formarse el conjunto de ecuaciones

$$x = f(\beta_0) + [\nabla^f(\beta_0)] [\Delta\beta]$$

y extraer el $\Delta\beta$ que mejor las satisfaga. Este es un problema de regresión lineal que se aborda formando las ecuaciones normales

$$[\nabla^{fT}(\beta_0) \nabla^f(\beta_0)] \Delta\beta = \nabla^{fT}(\beta_0) [x - f(\beta_0)] \quad (5.4)$$

de solución:

$$\Delta\beta = [\nabla^{fT}(\beta_0) \nabla^f(\beta_0)]^{-1} \nabla^{fT}(\beta_0) [x - f(\beta_0)] \quad (5.5)$$

Podemos corregir la estimación inicial β_0 mediante las soluciones proporcionadas por (5.5) de modo que una segunda estimación, más cercana a la óptima, viene dada por $\beta_1 = \beta_0 + \Delta\beta$. Empleando de modo iterativo este esquema, la secuencia

$$\beta_{n+1} = \beta_n + \lambda_n [\nabla^{fT}(\beta_n) \nabla^f(\beta_n)]^{-1} \nabla^{fT}(\beta_n) [x - f(\beta_n)] \quad (5.6)$$

converge hacia el óptimo local de forma asintótica. Este es el *algoritmo de Gauss o Gauss-Newton* ([3], [9]), que posee una tasa de convergencia cuadrática ([3]), lo que significa que $e_{n+1} = O(e_n^2)$, donde $e_n \equiv \|\beta_* - \beta_n\|$. Sin embargo, en cada iteración debe computarse la inversa de una matriz o bien una generalizada inversa.

En el caso de un sistema sobredeterminado, el algoritmo halla el valor de menor error cuadrático medio. Finalmente, si el número de incógnitas supera al de ecuaciones, la matriz $\nabla^T f(\beta_n) \nabla f(\beta_n)$ no tiene el rango completo, y la solución no es única. Como ya vimos, estos casos pueden ser englobados bajo el concepto de matriz pseudoinversa, de modo que (5.5) puede reescribirse como

$$\Delta\beta = \nabla^{f*}(\beta_n) [x - f(\beta_n)]$$

Esta es la aproximación adoptada por Fletcher en ([9]), donde propone un algoritmo recursivo de cómputo de la pseudoinversa en cada una de las iteraciones.

Si bien existen muchas variantes al algoritmo de Gauss-Newton ([10]), una muy utilizada en la práctica en tareas de espectroscopía como el ajuste de picos gamma ([2]), la determinación del espectro de líneas dada la matriz de transferencia del detector ([14]) y el análisis multivariado de espectros de absorción ([12]) es la de Levenberg-Marquardt ([2], [7], [13]). Es idéntico al de Gauss-Newton excepto en que mejora el índice de condicionamiento de $\nabla^T f(\beta_n) \nabla f(\beta_n)$ para mejorar la estabilidad del método:

$$\beta_{n+1} = \beta_n + [\mu_n I + \nabla^T f(\beta_n) \nabla f(\beta_n)]^{-1} \nabla^T f(\beta_n) [x - f(\beta_n)]$$

5.1.2.3 EL METODO DE NEWTON

Este es un método clásico de búsqueda de raíces de funciones. Como ilustra la Fig. 5.1b, los mínimos de la función del error cuadrático se encuentran en puntos donde se anula su derivada. Este es el fundamento del método de Newton aplicado a la

optimización de funciones. Observemos la Fig. 5.2. Estamos interesados en encontrar la raíz β_* de $F(\beta)$. La idea es tomar un valor inicial β_0 cercano a la misma y tratar de mejorarlo de forma progresiva. Tomando como valor inicial β_0 e incrementándolo en el valor definido por:

$$\Delta\beta_n = -\frac{F(\beta_n)}{F'(\beta_n)} \quad (5.7)$$

se obtiene una aproximación mejor β_1 . Si el valor inicial está suficientemente próximo al óptimo, la secuencia $\beta_0, \beta_1, \dots, \beta_n, \dots$ converge a este. Este procedimiento de aproximación se llama *método de Newton* ([5]). En el caso de que F fuese una función M -dimensional, el método de Newton se muestra como

$$\Delta\beta_{j_n} = -\frac{F(\beta_n)}{\nabla_j^F(\beta_n)}, \quad 1 \leq j \leq M. \quad (5.8)$$

El gradiente de la función $E(\beta)$ se anula en sus mínimos locales. Podemos, por tanto, determinar estos aplicando el método de Newton al sistema de M ecuaciones con M

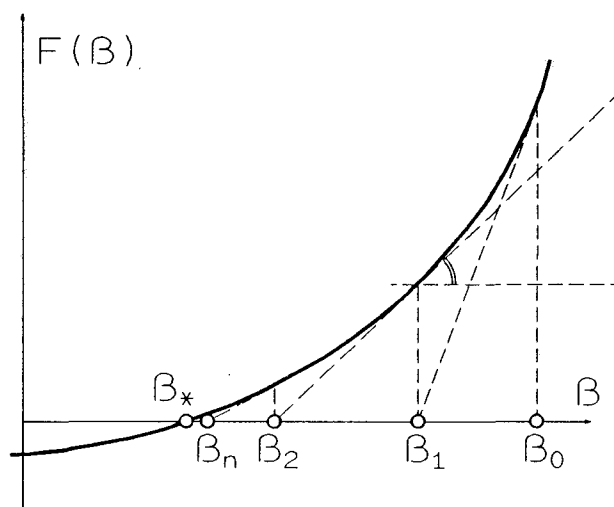


Fig. 5.2 Búsqueda local de la raíz β_* de una función $F(\beta)$ por medio del método de Newton o método de las tangentes.

incógnitas $\nabla^E(\beta) = 0$. Para descargar un tanto la notación sea $g(\beta) \equiv \nabla^E(\beta)$. El método de Newton aplicado a $g(\beta)$ resulta:

$$\Delta\beta_n = -\frac{g(\beta_n)}{\nabla^g(\beta_n)}$$

donde ∇^g es la matriz jacobiana de las M funciones de K parámetros que constituyen el gradiente de $g(\beta)$. $\nabla^g(\beta)$ se conoce como la matriz hessiana de la función $E(\beta)$. La matriz hessiana $H^F(\beta)$ de un función genérica $F(\beta)$ viene dada por $H_{ij}^F \equiv \delta^2 F / \delta\beta_i \delta\beta_j$. Particularizando la función F a la función de error cuadrático $E(\beta)$, la matriz hessiana adopta la expresión $H^E(\beta) = \nabla^{JT}(\beta)\nabla^J(\beta) + h^T(x - f(\beta))$, donde h es un vector N -dimensional cuyos elementos son matrices de dimensión $(K+2) \times (K+2)$, siendo el componente i -ésimo el hessiano de la función $f_i(\beta)$. El procedimiento de descenso es, así, el siguiente:

$$\Delta\beta_n = \lambda_n [\nabla^{JT}(\beta_n)\nabla^J(\beta_n) + h^T(x - f(\beta_n))]^{-1}(\beta_n) \nabla^{JT}(\beta_n) [x - f(\beta_n)]$$

Referencias a variantes de esta técnica pueden encontrarse en [8], entre ellas la de métrica variable. Esta es una de las técnicas de descenso más influyentes en espectroscopía gamma ([6], [7] y [8]). La aportación del *método de Métrica Variable* al método de Newton es el evitar el cálculo directo de la matriz $H^{E-1}(\beta_n)$ en cada iteración. En su lugar, es aproximada mediante otra matriz $L(\beta_n)$ de cálculo recursivo más sencillo ([6]). El precio pagado es una búsqueda del mínimo con mayor número de iteraciones pero con un esfuerzo de computación menor en cada una de ellas.

Los tres métodos expuestos guardan un gran parecido. Mostrados en conjunto:

Gradiente: $\Delta\beta_n = \lambda_n \nabla^{JT}(\beta_n) [x - f(\beta_n)]$

Gauss-Newton: $\Delta\beta_n = \lambda_n [\nabla^{JT}(\beta_n)\nabla^J(\beta_n)]^{-1}\nabla^{JT}(\beta_n) [x - f(\beta_n)]$

Newton:
$$\Delta\beta_n = \lambda_n [\nabla^{JT}(\beta_n)\nabla^J(\beta_n) + h^T(x - f(\beta_n))]^{-1} \nabla^{JT}(\beta_n) [x - f(\beta_n)],$$

puede detectarse con mayor facilidad el aumento de complejidad de la iteración con respecto a cada uno de ellos.

5.2 CRITERIOS METODOLOGICOS

Nuestra aproximación a la resolución del problema no lineal que representa el análisis de una mezcla será la de Gauss-Newton. Las razones las exponemos a continuación:

Primera. La ventaja de los métodos de gradiente frente a los de linealización de la función de ajuste es su campo de convergencia más amplio, pero su velocidad de aproximación al mínimo desciende rápidamente a medida que nos acercamos a éste. Además, su exactitud es menor que la de los métodos de linealización ([2]).

Segunda. Existen sensores de *espectroscopía de imagen* que proporcionan espectros de absorción por cada pixel de una escena bidimensional a tasas que aproximan el Gigabit/s ([11]). Esto significa que afrontar para cada espectro un esquema de descenso a partir del gradiente no es una vía lo suficientemente rápida para cubrir este volumen de información. Un operador de memoria asociativa que lleve a cabo el análisis de una secuencia rápida de espectros en simples operaciones de matriz-vector es una aproximación más adecuada.

Tercera. La realización en tecnología óptica de los algoritmos implicados resulta en principio viable ([1], [11]) y puede encontrar en este entorno computacional un campo abierto a la investigación.

5.3 EL ALGORITMO DE GAUSS-NEWTON EN EL PROBLEMA DE LA MEZCLA

En el anterior capítulo, vimos cómo descomponer espectros de radiación a través de una memoria asociativa cuando éstos estaban afectados por derivas instrumentales pequeñas respecto a las condiciones de calibración del instrumento. Sin embargo, las derivas pueden llegar a adquirir, con el paso del tiempo y la variación de las condiciones experimentales, magnitudes relativamente grandes que ponen en evidencia la insuficiencia de la expansión de Taylor alrededor de las condiciones experimentales en las que se capturaron los espectros del conjunto de referencia o, en términos más precisos, llevan a la dependencia no lineal del espectro problema respecto a estos. Ahora podemos entender el análisis de un espectro mezcla como un problema de estimación de un vector de parámetros lineales y un par de parámetros no lineales que caracterizan la distribución de las observaciones. En este apartado vamos a formular el problema en términos precisos dentro del marco de la teoría de optimización cuadrática no lineal examinada con anterioridad. Proponemos dos métodos de resolución del problema que derivan de dos modos de formular la función objetivo. Ambos serán examinados en sus aspectos computacionales.

5.3.1 UN PROBLEMA DE REGRESION NO LINEAL

Como ya hemos visto, es posible la aproximación al problema considerado este como de regresión no lineal. Sea el espectro problema x . Descomponer x supone ajustarlo al modelo de formación de mezcla sujeta a derivas instrumentales que viene dado por la función de canal f definida por:

$$f(x, c'^T, a, b) = \sum_{i=1}^K [c'_i r_i(ax + b)]$$

o, en forma vectorial

$$f(\mathbf{c}'^T, a, b) = \sum_{i=1}^K c'_i r_i(a, b) \quad (5.9)$$

y, por lo tanto, se trata de resolver el problema de

$$\underset{(c^T, a, b)}{\text{minimizar}} \quad \left\| x - \sum_{i=1}^K [c_i' r_i(a, b)] \right\|^2 \quad (5.10)$$

que puede ser abordado mediante el algoritmo de descenso de Gauss-Newton (5.6):

$$\beta_{n+1} = \beta_n + \nabla f^*(\beta_n) [x - f(\beta_n)] \quad (5.11)$$

En el caso que nos ocupa, el jacobiano de f , ∇f , es

$$\nabla^f = \left[\begin{array}{cccc} \frac{\delta f_1}{\delta c'_1}, & \dots, & \frac{\delta f_1}{\delta c'_K}, & \frac{\delta f_1}{\delta a}, & \frac{\delta f_1}{\delta b} \\ \frac{\delta f_2}{\delta c'_1}, & \dots, & \frac{\delta f_2}{\delta c'_K}, & \frac{\delta f_2}{\delta a}, & \frac{\delta f_2}{\delta b} \\ \vdots & & \vdots & & \vdots \\ \frac{\delta f_N}{\delta c'_1}, & \dots, & \frac{\delta f_N}{\delta c'_K}, & \frac{\delta f_N}{\delta a}, & \frac{\delta f_N}{\delta b} \end{array} \right],$$

cuyas $K+2$ columnas vienen dadas por:

$$\begin{aligned}\nabla_i^f &= r_i(a, b), \quad 1 \leq i \leq K, \\ \nabla_{K+1}^f &= \sum_{i=1}^K c_i' r_i'(a, b), \\ \nabla_{K+2j}^f &= \sum_{i=1}^K c_i' j r_{ij}'(a, b), \quad 0 \leq j \leq N-1, \end{aligned} \quad (5.12)$$

Si denominamos ${}^L\nabla^f = [\nabla_1^f \mid \dots \mid \nabla_K^f]$ y ${}^N\nabla^f = [\nabla_{K+1}^f \mid \nabla_{K+2}^f]$, entonces podemos escribir $\nabla^f = [{}^L\nabla^f \mid {}^N\nabla^f]$. Al pseudoinvertir ∇^f , obtenemos ∇^{f*} , que podemos expresar como $\nabla^{f*} = [[{}^L\nabla^{fp}]^T \mid [{}^N\nabla^{fp}]^T]^T$. La Fig. 5.3 clarifica esta notación.

De forma análoga, el vector de funciones $f(\beta)$ depende linealmente respecto de algunos de los parámetros y no linealmente respecto de otros ($\beta_i = c_i'$ para $1 \leq i \leq K$, $\beta_{K+1} = a$ y $\beta_{K+2} = b$), por lo que puede desdoblarse el vector β como $\beta^T = [{}^L\beta^T \mid {}^N\beta^T]$, donde ${}^L\beta$ es el vector de parámetros lineales y ${}^N\beta$ es el de los no lineales. Entonces el algoritmo de Gauss-Newton (5.11) se divide en dos procesos:

$${}^L\beta_{n+1} = {}^L\beta_n + {}^L\nabla^{fp}(\beta_n) [x - f(\beta_n)] \quad (5.13)$$

$${}^N\beta_{n+1} = {}^N\beta_n + {}^N\nabla^{fp}(\beta_n) [x - f(\beta_n)]$$

cuya interacción en cada iteración n es el cómputo de $f(\beta_n)$ y de $\nabla^{f*}(\beta_n)$. Ahora bien, sustituyendo la primera ecuación de (5.12) en (5.9) resulta que $f(\beta_n)$ es precisamente la combinación lineal ${}^L\nabla^f(\beta_n)c_n'$, por lo que

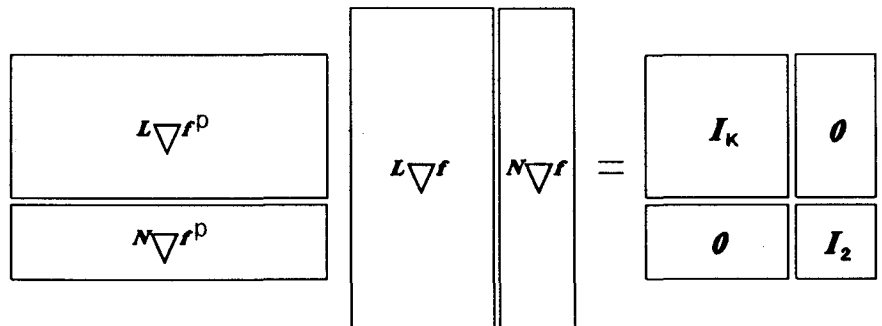


Fig. 5.3 Estructura y nomenclatura adoptada para la partición de la pseudoinversa del jacobiano de la función $f(\beta)$.

$$\begin{aligned}
 {}^L\nabla^{fp}(\beta_n)f(\beta_n) &= {}^L\nabla^{fp}(\beta_n){}^L\nabla^f(\beta_n)c_n \\
 &= I_K c_n \\
 &= {}^L\beta_n
 \end{aligned}$$

Por otra parte, sustituyendo la primera ecuación de (5.12) en ${}^N\nabla^{fp}(\beta_n)f(\beta_n)$ y dadas las relaciones ilustradas en Fig. 5.3, podemos comprobar que ${}^N\nabla^{fp}(\beta_n)f(\beta_n) = \mathbf{0}$, con lo cual el algoritmo (5.13) se simplifica, resultando finalmente:

$$\begin{aligned}
 {}^L\beta_{n+1} &= {}^L\nabla^{fp}(\beta_n)x \\
 {}^N\beta_{n+1} &= {}^N\beta_n + {}^N\nabla^{fp}(\beta_n)x
 \end{aligned} \tag{5.14}$$

Este algoritmo deja fijo el vector problema x , de parámetros desconocidos β_* y busca la estructura de cómputo ∇^f que es capaz de descomponerlo. En cada paso n es preciso computar la matriz $\nabla^f(\beta_n)$ y después pseudoinvertirla, por lo que el coste computacional es alto. Los algoritmos propuestos en el tercer capítulo pueden ser empleados a este fin.

5.3.2 UNA NUEVA FORMULACION DEL PROBLEMA

El método de Gauss-Newton (5.14) trata de buscar la configuración de parámetros ${}^N\beta_*$ tales que $\nabla^f({}^N\beta_*)$ sea el operador adecuado para descomponer la mezcla problema x . Sin embargo, podemos invertir esta idea de la manera siguiente: ya que x está posiblemente afectado por una deriva ${}^N\beta = [a, b]^T$, siempre es posible expresarlo como $x({}^N\beta)$ e introducir correcciones a la mezcla original $x({}^N\beta_0)$ de forma iterativa, de manera que los sucesivos $x({}^N\beta_n)$ converjan en un $x({}^N\beta_*)$ lo suficientemente cercano a una combinación lineal $f(c') = Rc'$ de los espectros de referencia r_1, \dots, r_K como para que pueda ser descompuesta

por un operador construido sobre ellos que no necesita ser continuamente reevaluado. El problema se plantea formalmente en los siguientes términos:

$$\underset{(c', a, b)}{\text{Minimizar}} \quad E(c', a, b) = \| x(a, b) - f(c') \|^2 \quad (5.15)$$

Definiendo $e(c', a, b) \equiv x(a, b) - f(c')$, ${}^L\beta = c$ y ${}^N\beta = [a, b]^T$, en lo que respecta al jacobiano de $e({}^L\beta, {}^N\beta)$, tenemos que

$$\frac{\delta e}{\delta {}^L\beta} = -\frac{\delta f}{\delta {}^L\beta}$$

y que

$$\frac{\delta e}{\delta {}^N\beta} = \frac{\delta x}{\delta {}^L\beta},$$

es decir, $\nabla^* = [-\nabla^f \mid \nabla^x]$. ∇^f es independiente del vector ${}^L\beta$ por ser f lineal sobre él, y su valor es la matriz R de columnas $r_1^{(0)}, \dots, r_K^{(0)}$. Por su parte, ∇^x sí depende de ${}^N\beta$. Es una matriz de dos columnas, $x'(a, b)$ y $jx'_j(a, b)$, $0 \leq j \leq N-1$, en adelante ${}^jx'(a, b)$. Bajo este planteamiento, podemos observar que ∇^* sólo depende de ${}^N\beta$, ya que tiene la forma $\nabla^*({}^N\beta) = [-R \mid x'({}^N\beta) \mid {}^jx'({}^N\beta)]$. El algoritmo de Gauss-Newton, aplicado a (5.15) se presenta como

$$\beta_{n+1} = \beta_n + \nabla^{*+} [x({}^N\beta_n) - f({}^L\beta_n)], \quad (5.16)$$

pero vamos a ver que se simplifica profundizando un poco más en el problema tal y como hicimos en el apartado anterior. Denotemos $\nabla^*({}^N\beta) = [-R \mid X({}^N\beta)]$ y dividamos su matriz pseudoinversa en dos submatrices de K y 2 filas respectivamente de modo que esta pueda ser expresada como $\nabla^{*+}({}^N\beta) = [R^{pT}({}^N\beta) \mid X^{pT}({}^N\beta)]^T$. Tenemos así que $-R^p({}^N\beta) R = I_K$ y que

$X^p(^N\beta) R = 0$, ya que $\nabla^{**} \nabla^* = I_{K+2}$. Podemos escribir entonces (5.16) como:

$$\begin{bmatrix} {}^L\beta_{n+1} \\ {}^N\beta_{n+1} \end{bmatrix} = \begin{bmatrix} {}^L\beta_n \\ {}^N\beta_n \end{bmatrix} + \begin{bmatrix} R^p(^N\beta_n) \\ X^p(^N\beta_n) \end{bmatrix} [x(^N\beta_n) - R^L\beta_n]$$

que resulta en

$${}^L\beta_{n+1} = R^p(^N\beta_n)x(^N\beta_n) \quad (5.17)$$

respecto a las contribuciones relativas, y

$${}^N\beta_{n+1} = {}^N\beta_n + X^p(^N\beta_n)x(^N\beta_n) \quad (5.18)$$

respecto a la deriva en ganancia y umbral. Podemos tomar como aproximación inicial ${}^N\beta_0$ el vector $[1, 0]^T$, que refleja la suposición de que la mezcla no esta afectada por derivas. Puede observarse que las estimaciones intermedias de las contribuciones no son necesarias, ya que ${}^L\beta$ no aparece en (5.18), y éstas se estiman una sola vez al aplicar (5.17) cuando se considera que $x(^N\beta_n)$ está libre de derivas instrumentales. En cada iteración n de (5.18) es necesario pseudoinvertir la matriz $[-R \mid x'(^N\beta_n) \mid {}^Jx'(^N\beta_n)]$, de N filas por $K+2$ columnas. De estas últimas, K invariantes con la iteración.

5.3.3 MEMORIA ASOCIATIVA EN EL ALGORITMO DE GAUSS-NEWTON

El algoritmo (5.18) puede modificarse de modo que el operador $X^p(^N\beta_n)$ no dependa de las mezcla problema del índice de la iteración n . El objetivo es conseguir sustituirlo por una aproximación *invariante con la iteración*, es decir, por una *memoria asociativa*. El operador de estas características que se propone es la matriz M que ya se

definió en el anterior capítulo para abordar el problema de la descomposición de la mezcla ante pequeñas derivas. Efectivamente, en la hipótesis de que cada referencia r_i está sometida a una deriva particular $(\Delta a_i, \Delta b_i)$, definimos

$$g(c', a, b) \equiv \sum_{i=1}^K c'_i r_i(a_i, b_i).$$

Así $\nabla^g(c', a, b)$, el jacobiano de $g(c', a, b)$, es una matriz de N filas por $3K$ columnas, que en las condiciones de referencia viene expresado como

$$\nabla_{3i-3}^g(c', 1, 0) = r_i,$$

$$\nabla_{3i-2,j}^g(c', 1, 0) = c'_i j r'_{ij}, \quad 0 \leq j \leq N-1,$$

$$\nabla_{3i-1}^g(c', 1, 0) = c'_i r'_i,$$

donde $1 \leq i \leq K$. A partir de $\nabla^g(c', a, b)$ se define $J = [J_0^T | J_1^T | J_2^T]^T$ del modo

$$J_0 \equiv [r_1, \dots, r_K],$$

$$J_1 \equiv \left[\frac{1}{c'_1} \nabla_1^g(c', 1, 0), \dots, \frac{1}{c'_K} \nabla_{3K-2}^g(c', 1, 0) \right],$$

$$J_2 \equiv \left[\frac{1}{c'_1} \nabla_2^g(c', 1, 0), \dots, \frac{1}{c'_K} \nabla_{3K-1}^g(c', 1, 0) \right].$$

La generalizada inversa de J puede ser expresada particionada en tres submatrices de K filas por N columnas como $J^{\star T} = [J_0^{pT} | J_1^{pT} | J_2^{pT}]$, de modo que definiendo $M \equiv J^{\star}$, se puede expresar $M^T = [M_0^T | M_1^T | M_2^T]$. Sean ς , γ y v tres vectores K -dimensionales. Tras la definición de estos últimos cuatro elementos se introduce el siguiente algoritmo de estimación de los parámetros de un espectro mezcla adquirido bajo condiciones de deriva instrumental:

$$E(\beta_0) = \infty$$

$$({}^N\beta_0) = [1, 0]^T$$

DO

{

$$\begin{bmatrix} \varsigma_n \\ \gamma_n \\ v_n \end{bmatrix} = \begin{bmatrix} M_0 \\ M_1 \\ M_2 \end{bmatrix} x({}^N\beta_n)$$

(5.19)

$$m = \left\{ \varsigma_{m_n} = \underset{k}{\text{Max}} \{ \varsigma_{k_n} \} \right\}, \quad 1 \leq k \leq K$$

$${}^N\beta_{n+1} = {}^N\beta_n + \frac{1}{\varsigma_{m_n}} \begin{bmatrix} \gamma_{m_n} \\ v_{m_n} \end{bmatrix}$$

} WHILE ($E(\beta_{n+1}) < E(\beta_n)$)

$${}^N\beta_* = {}^N\beta_n$$

$${}^L\beta_* = M_0 x({}^N\beta_*)$$

El principio de operación de este algoritmo es el siguiente: se presenta la mezcla x_n a la memoria y se obtiene el vector Mx_n . Sobre este resultado se opera una transformación $B(Mx_n)$ de selección de máximo y dos divisiones, de la que se obtiene una valoración de la deriva implicada $\Delta^N\beta_n = B(Mx_n)$. Con este dato se actúa sobre la mezcla introduciendo en ella la deriva estimada mediante una transformación $x_{n+1} = H((B \circ M)x_n, x_n)$ y el proceso continúa hasta que la función de error no disminuya. La Fig. 5.4 ilustra el algoritmo representando los bloques funcionales y el flujo de datos. En cada una de las iteraciones, la estimación de ${}^N\beta_n$ sigue en (5.19) el mismo criterio que en el modelo de pequeña deriva, es decir, se estima como la deriva asociada a la referencia de mayor contribución, es decir con mayor valor de ς_k .

Surge la cuestión de cómo decidir que β_n está suficientemente cercano a β_* como para darlo por bueno. Esta es una cuestión abierta en todos los métodos iterativos, por lo que la decisión debe hacerse "ad hoc". En este caso, la secuencia β_n lleva aparejada la secuencia monótonamente decreciente $E(\beta_n) \equiv \|R^L \beta_n - x^{(N)} \beta_n\|^2$. Cuando β_n está en una vecindad lo suficientemente pequeña se produce un fenómeno oscilatorio que rompe la tendencia a la baja del error, por lo que hemos adoptado como criterio de corte el estimar como β_* aquella β_n tal que $E(\beta_{n+1}) \geq E(\beta_n)$.

Como en el modelo de pequeña deriva, la prueba del encaje se realiza sobre $x^{(N)} \beta_*$ calculando $sen(\alpha)$. Este criterio puede fallar por dos razones: una deriva excesiva o bien un conjunto de referencia insuficiente. En este segundo caso puede aplicarse el algoritmo de selección de candidatos allí propuesto.

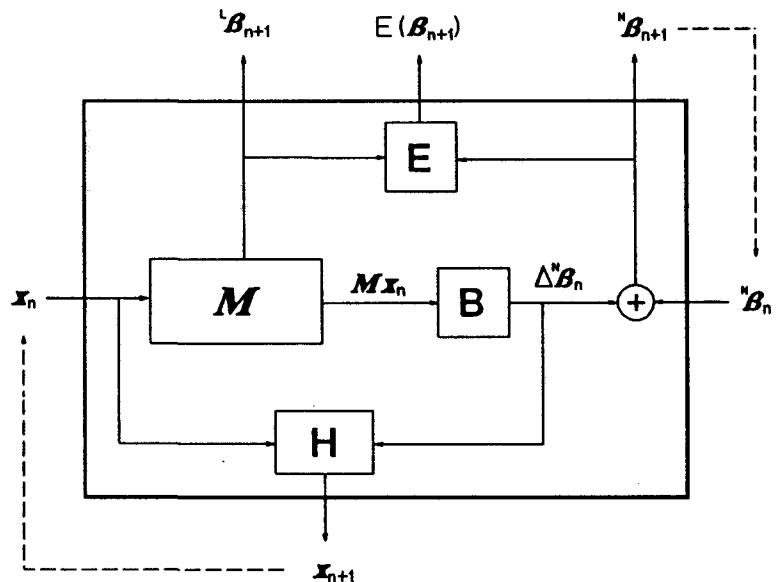


Fig. 5.4 El algoritmo (5.19). Representación esquemática de bloques funcionales y flujo de datos.

5.4 REGION DE CONVERGENCIA

La Fig. 5.5 muestra el valor de $\text{sen}(\alpha)$ obtenido al analizar el espectro de la 4.3. Como puede apreciarse, la memoria asociativa realimentada tolera un alto margen de distorsión en mezclas de un único componente. A su derecha se muestra de nuevo la 4.8, que representa el error obtenido por la memoria en ausencia de realimentación. En el dominio de $N\beta$, el campo de convergencia de la memoria parece ilimitado al analizar este caso. Sin embargo, cuando el número de componentes en la mezcla aumenta a k , el número de variables implicadas lo hace en $3k$. Se ha observado que este hecho provoca un fuerte impacto en la región de convergencia del algoritmo. A título de ejemplo, se tomó la mezcla de cuatro componentes (4.15), y se le sometió a un proceso de distorsión y posterior análisis de error mediante el algoritmo (5.19). Se utilizó el mismo conjunto de referencias de siete elementos, lo que representa la estimación de 21 variables. La Fig. 5.7 ilustra los resultados. La Fig. 4.21, correspondiente a la misma mezcla analizada mediante el modelo no realimentado, se muestra a su derecha a efectos comparativos. En la Fig. 5.7 se detectan claramente los límites de la región de convergencia. La experimentación con mezclas diversas muestra que estos límites, como cabía esperar, se estrechan a medida que aumenta el número de componentes.

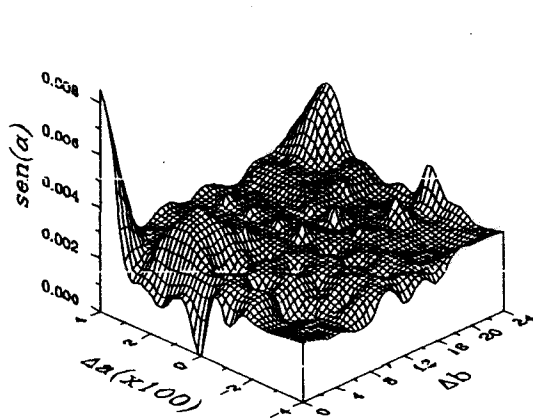


Fig. 5.5 BONDAD del ENCAJE como función de las derivas en ganancia y umbral presentes en el espectro de la Fig. 4.3. Memoria realimentada.

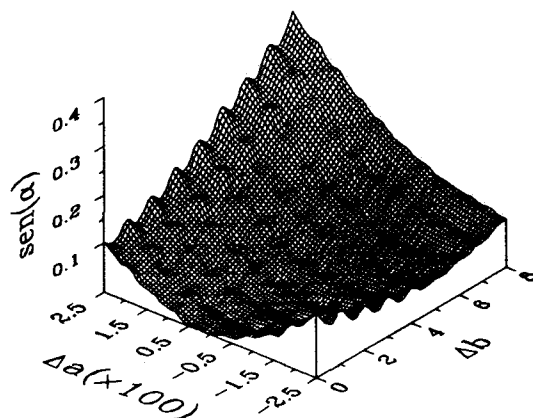


Fig. 5.6 BONDAD del ENCAJE como función de las derivas en ganancia presentes en el espectro de la Fig. 4.3. Memoria no realimentada.

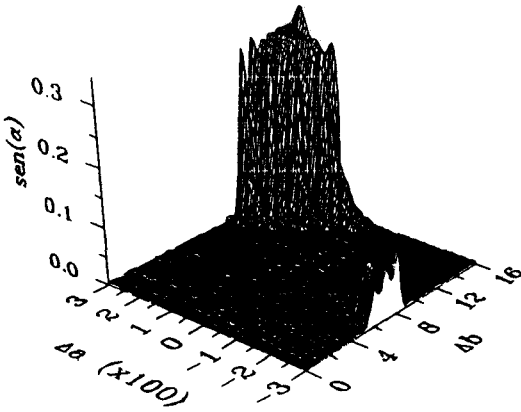


Fig. 5.7 BONDAD del ENCAJE de la mezcla (4.15) tras introducir en la misma derivas diversas. Memoria asociativa realimentada.

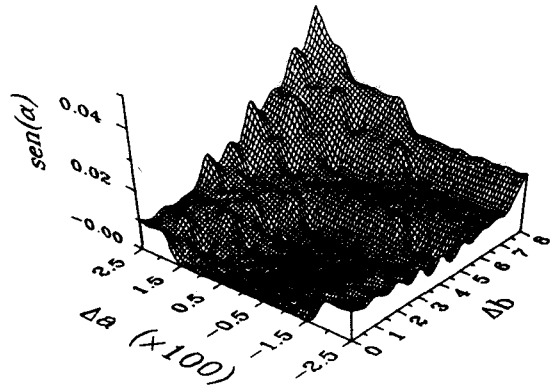


Fig. 5.8 BONDAD del ENCAJE de la mezcla (4.15) tras introducir en la misma derivas diversas. Seis términos de Taylor fueron considerados. Memoria no realimentada.

5.5 COMPUTO DE LAS TRANSFORMACIONES NO LINEALES

En el capítulo cuatro se trató del cómputo de derivadas a través de un *filtro lineal* que era una generalización del concepto de diferencias finitas. En el capítulo presente se ha considerado la posibilidad de la presencia de derivas severas en las mezclas problema, lo que hace necesario, como hemos estudiado, la necesidad de introducir cálculos relativos a transformaciones espectrales de acuerdo a incrementos en los parámetros de deriva. La orientación de la metodología de análisis hacia procesadores dedicados hace necesarios el diseño y la realización de una estructura de computación específica que lleve a cabo la *transformación no lineal* del vector N -dimensional $f(1, 0)$ en el vector con derivas $f(a, b)$ o, en notación continua, la función $f(x)$ en la función $h(x) \equiv f(ax+b)$. Esta es la tarea del bloque H en la Fig. 5.4, cuyo objetivo es la transformación de $f(x)$ (x_n) en $h(x)$ (x_{n+1}).

Si se define la función $g(x) \equiv f(ax)$, entonces $h(x) = g(x + b/a)$. Este encadenamiento de funciones posibilita el romper el proceso de transformación de $f(x)$ a $h(x)$ en

dos. Uno, de $f(x)$ a $g(x)$, y otro de $g(x)$ a $h(x)$. Ya que la segunda etapa no presenta dificultad por ser bien un retardo o bien un avance temporal, examinamos a continuación la primera. Expresemos el valor de ax en términos de sus partes entera y decimal de modo que $ax = E(ax) + D(ax)$. Si ax vale 137, entonces $f(ax)$ es conocido, pues es la componente 137-ava del vector f . Pero si ax vale por ejemplo 137'3, el valor de $f(ax)$ no está disponible y es necesario estimarlo. El método más intuitivo es la interpolación lineal entre los valores de $f(137)$ y $f(138)$. Aplicando esta técnica de aproximación:

$$\begin{aligned}
 g(x) &= f(ax) \\
 &= f(E(ax) + D(ax)) \\
 &\approx f(E(ax)) + D(ax) f'(E(ax)) \\
 &\approx f(E(ax)) + D(ax) [f(1 + E(ax)) - f(E(ax))]
 \end{aligned}$$

donde hemos estimado la derivada a partir de la primera diferencia de la función, si bien otros métodos más sofisticados pueden ser empleados, como ya conocemos. La fórmula anterior se empleará para calcular $g(0)$, $g(1)$, ..., $g(N-1)$, es decir, los todos los componentes del vector intermedio g excepto el último, al que se le asignará un cero o un valor igual al penúltimo:

$$\begin{aligned}
 g_i &\equiv f(E(ai)) + D(ai) [f(1 + E(ai)) - f(E(ai))], & 0 \leq i \leq N-2 \\
 g_{N-1} &\equiv 0
 \end{aligned} \tag{5.20}$$

La ecuación (5.20) da pie al diseño de un procesador sencillo de computo de g_i . Este se muestra en la Fig. 5.9. N ciclos del reloj completan la transformación. Este esquema, más un posterior desplazamiento de índice, conforman el bloque funcional H del la Fig. 5.4.

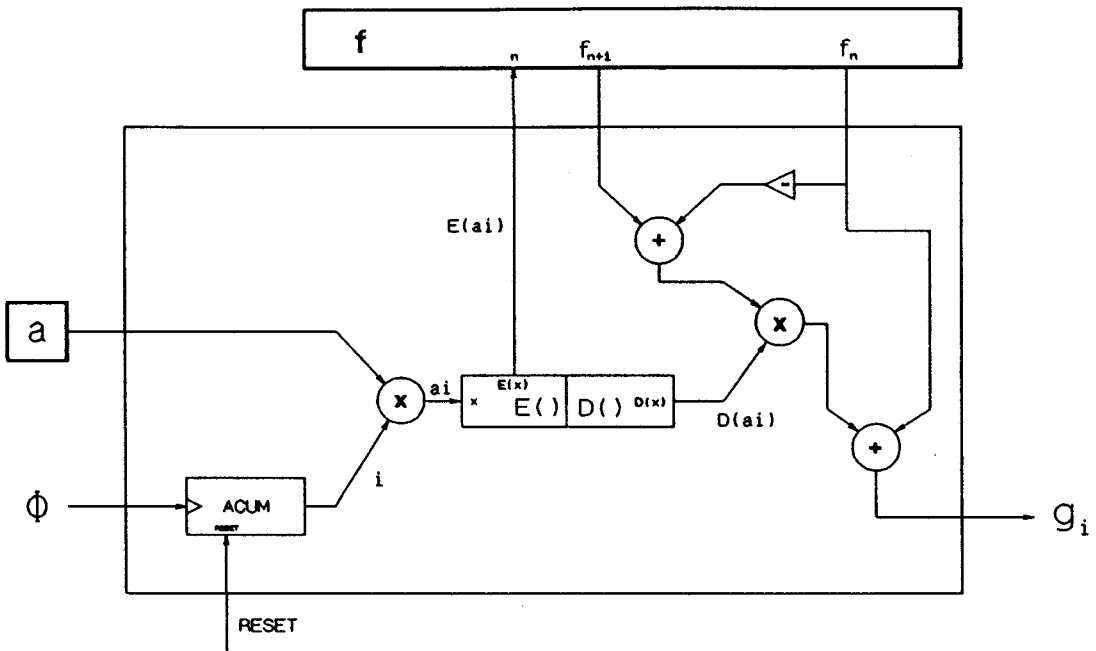


Fig. 5.9 Diagrama esquemático de una estructura de computación que transforma $f(x)$ en $g(x)=f(ax)$.

5.6 BIBLIOGRAFIA

- [1] Barnard, B. *et* Casasent, D. P., "Optical Neural Net For Classifying Imaging Spectrometer Data", *Applied Optics*, Vol. 28, No. 15, August, (1989), pp. 3129-3133.
- [2] Bevington, P. R., "Data Reduction and Error Analysis in the Physical Sciences", McGraw-Hill, (1969).
- [3] Brown, K. M. *et* Dennis, J. E., "Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximation", *Numer. Math*, 18, (1972), p. 289-297.
- [4] Davidon, W. C., "Variable Metric Method for Minimization", Argonne National Laboratory, Report ANL-5990 Rev., (1959).

- [5] Demidovich, B. P., *et* Maron, I. A., "Cálculo numérico fundamental", Editorial Paraninfo, (1985).
- [6] Fletcher, R., "A New Approach to Variable Metric Algorithms", *Computer J.*, 13, (1970), pp. 317-322.
- [7] Fletcher, R., "A Modified Marquard Subroutine For Non-Linear Least Squares", Report AERE Report R-6799, Theoretical Physics Division, Atomic Energy Research Establishment Harwell, Berkshire, (1971).
- [8] Fletcher, R. *et* Powell, M. J. D., "A Rapidly Convergent Descent Method For Minimization", *The Computer Journal*, Vol. 6, (1963), pp. 163-168.
- [9] Fletcher, R., "Generalized Inverse Method for the Best Least Squares Solution Of Systems of Non-Linear Equations", *The Computer Journal*, No. 10, (1968), pp. 392-399.
- [10] Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least-Squares", *Technometrics*, Vol. 3, No. 2, May, (1961), pp. 269-280.
- [11] Liu *et* Casasent, "Optical Processing of Imaging Spectrometer Data", *SPIE*, Vol. 938, (1988).
- [12] Maeder, M. *et* Zuberbuler, A. D., "Nonlinear Least Squares Fitting of Multivariate Absortion Data", *Analytical Chemistry*, 62, (1990), pp. 2220-2224.
- [13] Marquard, D. W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", *SIAM Journal on Applied Mathematics*, Vol. 11, No. 2, June, (1963). pp. 431-441.
- [14] Sekimoto, H., "An Unfolding Method Leading to a Positive Solution Only", *Nuclear Instruments and Methods*, 288, (1984), pp. 129-132.

Discusión de resultados

El punto de contraste crítico de todo trabajo de investigación es sin duda el dedicado a comentar los resultados obtenidos a lo largo del mismo. Se debe realizar una valoración cuidadosa de todos y cada uno de los aspectos de relevancia conseguidos en el desarrollo del trabajo de investigación, resultando esta tarea especialmente dura para el autor, ya que es éste quien con menos objetividad puede enjuiciar su propio trabajo, siendo la comunidad científica la que realmente está llamada a emitir opiniones sobre el mismo. Este apartado, pretende, no obstante, ser una crítica objetiva del autor sobre su propio trabajo. Es difícil que a lo largo de la misma no se incurra en sesgos, comprensibles por otro lado, de destacar aspectos menos relevantes que otros, pero con los que el autor se ha podido ligar de una forma más afectiva, en base a que le hayan podido aportar más satisfacciones desde el punto de vista de estudio y elaboración, que otros que, aun de mayor transcendencia, hayan sido personalmente menos agradecidos. Estos factores pueden distorsionar una labor de autocrítica, como la que se pretende realizar en este punto y, por ello, el autor ruega la comprensión del lector si acaso se apreciase dicho efecto en la discusión que sigue.

Como principal metodología para enfocar la citada discusión se ha adoptado la de contrastar los resultados obtenidos con los objetivos que inicialmente se habían planteado,

a los que denominaremos resultados afines a los objetivos, expresando el grado de consecución de los mismos, e ilustrando aquellos resultados obtenidos que no se hallen directamente relacionados con ninguno de dichos objetivos, como lo que podríamos llamar resultados colaterales. Pasaremos a ocuparnos adecuadamente de unos y otros.

Como primer resultado afín al objetivo específico primero se destaca que la metodología de regresión lineal múltiple utilizada es perfectamente válida para realizar la identificación de presencia de la actividad de un radionúclido incluido en la base de referencia, señalando de forma inequívoca dicha presencia en tal caso, mediante la formación de un peso no negativo en el vector de contribuciones. En el caso de que dicho radionúclido no se halle presente, el valor del peso asociado al mismo se halla siempre por debajo de un cierto umbral próximo a cero. Númerosos análisis de espectros compuestos, algunos de ellos documentados en las publicaciones [33] y [34] del capítulo tres, constituyen la probatoria de todos los extremos afirmados en el presente punto.

Como resultado colateral al recientemente expuesto, cabe mencionar que en el caso de que en un espectro compuesto exista presencia de actividad de uno o varios radionúclidos no incluidos en la base de referencia, se ha propuesto el uso de un coeficiente de fiabilidad de bajo coste computacional, que permite poner en evidencia dichos casos, para que los mismos sean estudiados utilizando otras bases de referencia, o bien para que se amplíen convenientemente éstas. Dichos extremos se hallan estudiados y documentados en el sexto apartado del tercer capítulo.

Como segundo resultado afín con el segundo objetivo específico, es de destacar que el método desarrollado, mediante la construcción del denominado vector de contribuciones, permite determinar la cuantía relativa precisa en la que un determinado elemento presente en la base de referencia contribuye a la formación del espectro compuesto en estudio. Este proceso de cuantificación resulta altamente preciso si no se producen derivas de consideración en los procesos de construcción espectral referidos. Estos aspectos, estrictamente ligados al de la presencia de un elemento en la mezcla, son tratados en las referencias citadas y en los apartados 3.1 y 3.6.2 del tercer capítulo.

En relación con el tercer objetivo específico, cabe citar que los métodos

anteriormente mencionados permiten la inclusión demostrada de hasta 30 espectros individuales en la base de referencia, sin que se produzcan distorsiones o errores inaceptables en el vector de contribuciones, validando en esta forma la metodología propuesta. Los procedimientos de construcción de la pseudoinversa son sensibles al aumento de componentes individuales en la base de referencias hasta una cierta extensión, que cubre la mayoría de los casos que se pueden dar en un trabajo de laboratorio o de campo. Este no es el caso de una central nuclear, donde el número de radio-núclidos individuales que contribuyen a la formación de un espectro puede ser muy alto, como se comenta en el apartado de Conclusiones. En tales casos, la metodología desarrollada en el presente trabajo habría de ser sustancialmente remodelada, para admitir su extensión a costa de sacrificar otros aspectos relacionados, tales como la precisión o la robustez frente a derivas, de la que se tratará en breve. Los aspectos relacionados con la dimensionalidad máxima de la base de referencias se tratan en el apartado 3.6.2 desde un punto de vista teórico en lo que se refiere a la fiabilidad del análisis. La confirmación experimental de la degradación que sufre el vector de contribuciones cuando la base de referencias contiene más de las necesarias para descomponer la mezcla problema puso de manifiesto la necesidad de disponer de una justificación teórica de este efecto, así como de un algoritmo de refinamiento de la base de bajos requisitos en cuanto a aritmética. Este algoritmo se propone y valida en el apartado 3.7.

Por lo que respecta al cuarto objetivo específico, se puede hacer hincapié en que una característica fundamental que se ha buscado en los algoritmos de construcción de pseudoinversas ha sido la de su naturaleza recursiva y fácilmente ampliable, de modo que tanto los métodos basados en el algoritmo de Pyle, como los centrados en torno al algoritmo de Widrow-Hoff presentan dicha característica. En el caso del algoritmo de Pyle, el proceso constructivo para aumentar la dimensionalidad del resolutor de orden k al de orden $k+1$ resulta notoriamente sencillo, de manera que dicha actualización podría incluso llevarse a cabo "on line", sobre procesadores específicos de complejidad media. La justificación de estos extremos puede verse en los apartados 3.2 y 3.3.

Como resultado colateral con dicho objetivo, cabe citar que se ha determinado que el procedimiento adecuado para disponer de métodos que admitan una ampliación recursiva es el de construcción de la pseudoinversa, ya que construida esta estructura, se dispone de un resolutor para un conjunto de problemas comunes o afines y por ello se han descartado

otras metodologías que permitieran la construcción recursiva del vector de contribuciones.

Por lo que respecta al quinto objetivo específico, cabe señalar que los resultados que se documentan en la referencia [33] del tercer capítulo garantizan que en espectros compuestos producidos por un número bajo de radio-núclidos se puede diferenciar o resolver adecuadamente con proporciones muy dispares de componentes. Típicamente, se puede resolver en proporciones de hasta 1:100 o incluso aún más desproporcionadas, cuando el espectro compuesto está formado por dos radio-núclidos. Hay que decir, sin embargo, que esta capacidad de resolución entre componentes disminuye al aumentar el número de componentes en la mezcla y la inestabilidad de las condiciones ambientales durante el periodo de captura de la misma.

Los resultados afines al sexto objetivo específico son posiblemente los más espectaculares de los producidos en relación con el presente trabajo. Es de destacar que el problema de las distorsiones en la formación de espectros de radiación a resultas de las derivas en ganancia y "offset" de la instrumentación involucrada han planteado serios problemas a la posible automatización de los procedimientos de análisis espectrográfico de espectro completo. Uno de los principales resultados obtenidos en el presente trabajo en este sentido ha sido el de establecer un modelo de distorsión del espectro compuesto, que se ha demostrado altamente eficaz en el tratamiento del problema. Estos aspectos se documentan en detalle en el apartado 4.1.

Un primer resultado colateral al citado objetivo, que es asimismo destacable, es el de haber conseguido que el modelo de distorsión espectral utilizado sea perfectamente encajable dentro del mismo Modelo de Regresión Lineal Múltiple desarrollado para la construcción de los resolutores, con lo que se ha unificado la resolución de los casos con derivas con aquellos en los que éstas no estén presentes. El método de ampliación del resolutor, utilizado para representar estos casos, se describe y discute en el apartado 4.2.

Un segundo resultado colateral con el objetivo antes mencionado es la confirmación experimental de que la memoria asociativa, una vez ampliada y a través de un proceso de realimentación inspirado en los métodos de optimización no lineal, es capaz de resolver mezclas con fuerte deriva. El desarrollo de este extremo es el objeto del quinto capítulo.

Finalmente, en relación con el séptimo objetivo específico, cabe mencionar que se ha puesto el énfasis en la determinación de algoritmos de bajo coste computacional, y de requisitos aritméticos sencillos. En este sentido, el algoritmo de Pyle presenta un coste computacional razonablemente bajo, aunque demanda de operaciones de división, mientras que su correlato de Widrow-Hoff exige un mayor número de operaciones, pero a cambio precisa únicamente de operaciones de producto interno (sumas y multiplicaciones). Un reto pendiente, que constituye una de las principales líneas de continuación del presente trabajo, consiste en el estudio de las posibilidades de implementación de esta última técnica con aritmética restringida y el de la sensibilidad de dicha metodología ante la precisión de la representación utilizada. En este contexto se están iniciando otros trabajos relacionados que pueden aportar resultados altamente esperanzadores. Finalmente, cabe citar que el proceso de diseño de un procesador dedicado que permita soportar estos algoritmos en forma autónoma constituye uno de los desafíos más interesantes que este trabajo deja planteados sobre el tapete.

Conclusiones

El trabajo desarrollado constituye un estudio sobre la aplicación y la posterior implementación VLSI del modelo estadístico de regresión múltiple con el fin de resolver un problema concreto que es el análisis de espectros de radiación gamma de baja resolución. La adopción de esta metodología de descomposición de espectro completo se sustenta sobre dos argumentos fundamentales. Primero, desde la perspectiva más amplia de la espectroscopía nuclear, la técnica alternativa de búsqueda de fotopicos en el espectro problema no resulta apropiada en espectros de baja resolución, debido a la dificultad o imposibilidad de discriminar entre picos anchos y solapados entre sí. Segundo, el estudio se enmarca dentro de un proyecto más amplio que es la construcción de una gamma-cámara que incorpore la formación selectiva de la imagen de la radiación procedente de un radioisótopo determinado. La consiguiente necesidad de rapidez en el análisis del espectro correspondiente a cada uno de los pixel de la imagen parece desaconsejar la búsqueda de fotopicos, por constituir el ajuste de cada uno de ellos un problema adicional de regresión con varios parámetros no lineales. Recientes publicaciones sobre espectroscopía de absorción adoptan la vía de análisis de espectro completo.

La técnica de regresión del espectro completo posee, sin embargo, una debilidad fundamental, ya que exige el conocimiento de todos y cada uno de espectros de los

radionúclidos que constituyen la fuente de radiación. Este requisito resulta a veces difícil de satisfacer por la dificultad de aislar químicamente algunos radionúclidos determinados. Este fue un comentario muy útil que se hizo al autor en su visita a la Central Nuclear de Vandellós II (Tarragona) en 1992. No existe solución al problema de la falta de referencias y este hecho debe entenderse como una restricción en la aplicación del método. En las centrales nucleares es común encontrar hasta cien isótopos distintos en un espectro problema. Por supuesto, estos son espectros capturados con detectores de Ge(Li) y muestran picos muy bien definidos, que no es el objeto del presente trabajo.

Las dificultades de aplicación encontradas en el ámbito de las centrales nucleares no pueden restar generalidad a la aproximación seguida en el trabajo. Al contrario, tiene aplicación allá donde se presente un problema de regresión, sujeto o no a deriva instrumental, que en adelante denominaremos el problema de la mezcla para subrayar la generalidad aludida. A lo largo de la exposición se ha mostrado cómo este problema puede ser resuelto de forma rápida y elegante a través de un operador lineal denominado memoria asociativa.

El objetivo de implementación de algoritmos sobre procesadores dedicados se interpreta en el trabajo como la implementación de la memoria misma así como los algoritmos que la construyen. Cada fila de la memoria se considera como un procesador sencillo de producto interno cuyos pesos son los componentes de la fila. Una memoria de $3 \times M$ filas es implementada como un conjunto de $3 \times M$ procesadores. El análisis de un espectro de radiación se lleva a cabo con $3 \times M$ productos internos que pueden desarrollarse en paralelo.

El tamaño de la memoria es un aspecto importante que acomete el trabajo, ya que la aportación de demasiadas referencias a la formación de la misma degrada la calidad de la descomposición, es decir, pequeñas distorsiones aleatorias en la mezcla provocan que la memoria pueda llegar a detectar contribuciones de referencias no presentes. Este efecto no es más que la manifestación de la inevitable pérdida de estabilidad que se produce en todo problema inverso en el que crece el número de variables a determinar. La introducción de términos de regularización en la función objetivo es una forma generalmente utilizada de

paliar este problema, si bien introduce un componente artificial en la solución difícilmente aceptable en la aplicación tratada. En su lugar, se propone construir la memoria de una forma adaptativa, teniendo en cuenta la mezcla problema a la hora de introducir una nueva referencia en la memoria siguiendo una pauta de correlación mezcla-referencia. Un algoritmo bien definido formaliza esta aproximación.

La construcción adaptativa de la memoria requiere de un algoritmo de actualización de pesos cuando una nueva referencia es incorporada a la misma. Se responde a este punto con la propuesta de dos algoritmos cuya característica común es que ninguno de ellos requiere de la operación de raíz cuadrada: el algoritmo de Greville y el algoritmo de Widrow-Hoff. Este particular cobra especial importancia porque conduce a una mayor simplicidad en el diseño de la unidad que control del algoritmo y a una disminución de los tiempos de cómputo. El algoritmo de Greville requiere de la operación de división. El algoritmo de Widrow-Hoff únicamente necesita de los operadores de suma y multiplicación. El peso concedido a la simplicidad de la aritmética y con ella al área de silicio del diseño global, el que se otorga a la velocidad de proceso y el dado a la precisión de la memoria establecen el equilibrio de fuerzas que compromete la arquitectura del sistema final. Las simulaciones llevadas a cabo muestran el comportamiento de los dos algoritmos bajo precisiones diversas de representación numérica poniendo de manifiesto la viabilidad de ambas aproximaciones.

La mezcla problema se presenta más de una vez a la memoria en el caso de derivas severas. Este es el mecanismo que se propone como alternativa al de Gauss-Newton, mucho más lento, ya que necesita computar una matriz pseudoinversa en cada iteración. A su favor, el método de Newton ofrece una tolerancia a la deriva mayor que la memoria con realimentación. Este hecho se hace patente a medida que aumenta el número de componentes en la mezcla problema y la memoria crece en su número de filas. Velocidad frente a tolerancia a derivas constituye por tanto el compromiso entre ambos métodos. En cualquier caso, el cómputo de las generalizadas inversas constituye la función específica de los algoritmos de Greville y Widrow-Hoff. En consecuencia, ambos métodos pueden ser implementados sobre la misma arquitectura, salvo el diseño de la unidad de control.

Líneas Futuras

Cuando un trabajo como el presente inicia una línea de interés para un grupo, en la que no se había trabajado con anterioridad, existe la impresión de que se abren muchas líneas asociadas como continuación del mismo, si bien el tiempo se encarga posteriormente de determinar cuales de aquellas constituyen verdaderas fuentes productivas. Por ello, este apartado no quiere incurrir en demasiadas pretensiones y tratará de evitar citar aquellos aspectos del trabajo realizado que tienen menos visos de continuidad en el marco actual del Grupo de Tecnología de Computadores o, en todo caso, se hará únicamente una breve mención a los mismos. Hecha esta salvedad, el resumen de las líneas de trabajo que se pueden haber abierto mediante la presente tesis es el siguiente:

1. Como ya apuntamos en el apartado de conclusiones, el tamaño de la memoria es un aspecto importante en el estudio que acomete el presente trabajo, ya que la aportación de demasiadas referencias a la formación de la misma hace inestable el resultado de la descomposición ante pequeñas distorsiones aleatorias de la mezcla. La introducción de términos de regularización en la función objetivo, como la minimización de la norma del vector de contribuciones, sería una forma de paliar este problema. Esta técnica, sin embargo, introduce una componente artificial en la solución que es difícilmente asumible en la aplicación tratada. No obstante, otras

aplicaciones menos exigentes en este aspecto podrían beneficiarse de ella, por lo que la determinación de algoritmos adaptativos que la contemplen presenta un indudable interés.

2. La construcción adaptativa de la memoria requiere de un algoritmo de actualización de pesos cuando se incorpora una nueva referencia a la misma. A lo largo del trabajo se ha propuesto el de Widrow-Hoff en base a que únicamente requiere de operaciones de suma y multiplicación. El peso concedido a la simplicidad de la aritmética, y con ella a la reducción de coste en área de silicio en el diseño global, frente a la precisión de la memoria, constituyen los factores a tener en cuenta para alcanzar un compromiso final. El estudio puede continuarse por medio de simulaciones que muestren el comportamiento de los algoritmos involucrados bajo precisiones diversas, y con diferentes representaciones numéricas, poniendo de manifiesto la viabilidad de las mismas. Esta línea se halla en estudio en el momento actual por medio de otro trabajo que pretende elevarse a Tesis Doctoral, siendo los resultados obtenidos hasta el momento altamente prometedores.
3. Estudio de la disparidad en las contribuciones, y su influencia en la precisión de los algoritmos de detección.
4. Diagnóstico automático del sistema de detección ante derivas en su función de transferencia.
5. Implementación de procesadores dedicados para la detección y cuantificación de espectros de radiación, utilizando como base el operador de matriz pseudoinversa.
6. Estudio de los aspectos aritméticos de dichos procesadores, para la optimización y minimización de los requisitos de los mismos.
7. Extensión de los métodos elaborados aquí para el tratamiento de espectros

compuestos en el margen de infrarrojos, visible o ultravioleta. En relación con este último punto, cabe decir que el mismo permitiría ampliar notablemente el campo de incidencia de la metodología desarrollada, al permitir su extensión a un dominio de aplicaciones mucho más rico, como puede ser en espectroscopía óptica, colorimetría, percepción del color, análisis de pigmentos fotoemisores y fotoabsorbentes, coloreado de imágenes, detección y gestión remota de recursos por fotografía aérea o por satélite, tratamiento y restauración de obras de arte, etc. En este sentido se está configurando una línea de trabajo que abarcará investigadores de varias Universidades nacionales.

Agradecimientos

Quiero expresar mi gratitud a todos aquellos que, de una manera u otra, me han brindado la posibilidad de acometer este trabajo y de llevarlo a buen término. Una buena parte de mi esfuerzo ha sido fruto del afán por no defraudarles. El apoyo, la comprensión y el cariño de mis padres, Marce y Jacinto, de mi hermano Pedro Pablo y de mi prima Feli han hecho que todo fuese más fácil. Pedro Gómez, que ya fue tutor de mi trabajo de fin de carrera, me propuso en su día el proyecto más ambicioso de realizar una tesis doctoral. A veces, las personas actúan movidas por la confianza que les inspiran otras. He de reconocer que en un primer momento no fue el entusiasmo por el tema sino el mismo hecho de trabajar junto a Pedro Gómez y Victoria Rodellar lo que me movió a aceptar el reto. Mi relación personal con ellos ha sido siempre entrañable. Desde aquí quiero corresponder a todo lo que, en el sentido más amplio, he aprendido a su lado y al amable trato que, desde su sencillez, siempre me han dispensado.

Han sido muchas las horas pasadas en el laboratorio del Departamento de Arquitectura y Tecnología de Sistemas Informáticos y no es posible dejar de mencionar a todos aquellos que las hicieron más llevaderas. Mariano y Marisa, Charly, Alfonso Moreno, Fernando Naharro, Consuelo Gonzalo, Víctor Nieto, Antonio Díaz y Virginia Peinado, Elvira Martínez, Margarita Pérez, Sol Martín y, muy especialmente, Chema Gallardo,

Manolo Sánchez y Alfonso Castillo por su generosidad y por su amistad. Consuelo Gonzalo se tomó la paciencia de revisar los primeros manuscritos. Sus comentarios, siempre certeros, contribuyeron notablemente a mejorar su estilo. A todos vosotros, os manifiesto mi más sincera gratitud y estima personal.

La otra gran parte del tiempo dedicado a este trabajo ha transcurrido en el CIEMAT. Junto con Germán García, José Manuel Pérez, Antonio Bru, Juan Carlos Oller, Pablo Aguayo, Jesús Marín, Luis Amigo y Nacho Carvallo gané el trofeo a la deportividad en la liga de futbito 91/92. José Luis de Pablos sabe de amplificadores y casas de madera y no practica deporte alguno. Pablo Aguayo, Luis Amigo, Nacho Carvallo y David Francia, componentes del equipo de Microelectrónica, "El Meka", son los compañeros más ingeniosos, alegres y divertidos que a uno pueden tocarle en suerte. De verdad, muchas gracias. Me he divertido con vosotros.